



HAL
open science

La genèse de l'élision

Timothée Premat

► **To cite this version:**

Timothée Premat. La genèse de l'élision. Linguistique. Université Côte d'Azur, 2023. Français.
NNT: 2023COAZ2051 . tel-04574584

HAL Id: tel-04574584

<https://theses.hal.science/tel-04574584>

Submitted on 14 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ
CÔTE D'AZUR

ÉCOLE DOCTORALE
SOCIÉTÉS, HUMANITÉS,
ARTS ET LETTRES

THÈSE DE DOCTORAT

La genèse de l'élision

Timothée PREMAT

Base, Corpus, Langage (UMR 7320)

**Présentée en vue de l'obtention
du grade de docteur en Sciences du
Langage**

d'Université Côte d'Azur

Dirigée par : Tobias SCHEER

Soutenue le : 11/12/2023

Devant le jury, composé de :

Joaquim BRANDÃO DE CARVALHO,
Professeur, Université Paris 8

Céline GUILLOT-BARBANCE, Maîtresse de
conférences HDR, ENS de Lyon

Haike JACOBS, Professor, Radboud
Universiteit

Thomas RAINSFORD, Akademischer Rat,
Universität Stuttgart

Tobias SCHEER, Directeur de Recherche,
CNRS (UMR 7320), directeur de la thèse

La genèse de l'élision

Jury :

Président du jury

Haike JACOBS, Professor, Radboud Universiteit

Rapporteurs

Céline GUILLOT-BARBANCE, Maîtresse de Conférences habilitée à diriger des recherches, ENS de Lyon

Haike JACOBS, Professor, Radboud Universiteit

Examineurs

Joaquim BRANDÃO DE CARVALHO, Professeur des Universités, Université Paris 8

Thomas RAINSFORD, Akademischer Rat, Universität Stuttgart

Directeur de thèse

Tobias SCHEER, Directeur de Recherche, CNRS (UMR 7320)

La genèse de l'élision

Cette thèse traite de l'origine de l'élision des voyelles finales atones en ancien français. Elle s'appuie sur des corpus numériques (*Nouveau Corpus d'Amsterdam* et *Base de Français Médiéval*), analysés via deux programmes informatiques développés pour l'occasion, Comparalem et le PAM (Programme d'Analyse Métrique). Le premier programme permet la détection semi-automatique des lemmes sujets à des suppressions de -e final, tandis que le second permet l'annotation et l'analyse automatique des propriétés métriques de textes versifiés.

L'omission graphique de voyelles finales atones ne concerne que certains monosyllabes fonctionnels, ce qui correspond à la première période de la phonologie médiévale, où le domaine computationnel ne peut contenir qu'un seul mot lexical. En revanche, en métrique, toute voyelle finale atone est élidée au sein d'un hémistiche ou d'un vers, ce qui est cohérent avec la seconde période de la phonologie médiévale, où le domaine computationnel s'élargit et peut contenir plus d'un mot lexical. Nous résolvons cette contradiction entre les données graphiques et les données métriques en avançant que, puisque la métrique n'a pas de constituant plus petit que l'hémistiche, elle n'est pas un témoin fiable de la première phase de la phonologie de l'ancien français, tandis qu'elle révèle au contraire des propriétés qui, dans la seconde phase de l'ancien français, sont cachées sous les habitudes graphiques des scribes, qui continuent à écrire des voyelles pourtant élidées. L'analyse formelle que nous proposons parvient à modéliser ces deux états de la phonologie médiévale en utilisant les mêmes représentations et processus, et démontre qu'il n'est pas nécessaire de postuler une motivation anti-hiatique à l'élision.

Mots clefs : Phonologie, diachronie, français, ancien français, TAL.

Genesis of Elision in Medieval French

This thesis focuses on the origin of elision (stressless final vowel deletion before a vowel-initial word) in Old French. It is based on two digital corpora, the *Nouveau Corpus d'Amsterdam* and the *Base de Français Médiéval*. The data is treated through two computer programs I developed, Comparalem and the PAM. The former allows for semi-automatic detection of lemmas with deleted final vowel, statistical computation, and mapping of data in time and space; the later performs automatic prosodic tagging and metrical analysis of versified texts.

Graphic omission of final stressless vowels, which only affects monosyllabic function words, matches what is expected from the first stage of Medieval French phonology, in which computational domain is restricted to one lexical word; hence elision only affects clitics. On the opposite, in metrics, almost all final stressless vowel is elided inside a hemistich or a line; this is consistent with the second phase of Medieval French phonology, where computational domain comprises one or several lexical words. I argue that, because metrical paraphonology does not know any domain smaller than the hemistich, metrical testimonies are irrelevant for the phonology of the first stage of Old French, while it reveals properties that are hidden in the graphic habits of scribes in the second stage of Old French phonology. I propose a formal analysis that derives these two stages of elision using the same representations and tools, and argue that there is no need for a hiatus reducing rule.

Keywords : Phonology, diachrony, French, Old French, NLP.

A O' LĚCTEVRS

COmbien qe d'une pouure confide-
raçion la pluſgran' partie de no'Frā-
çoēs foēt en fantazie qe la pouurfuyte
d'une grammere foēt trop diffiçil' e
pre'q'impofsibl' en noſtre lange : je n'en n'ey
pas pourtant ſi dezefperé qe je n'aye fet qelqe
dilijençe d'en çerçher qelqes moyēs , e reğles.

— Louiſ Meigrêt,
Le treçté de la GRAMMERE FRANÇOËZE,
Fol. 2v

Sommaire

Sommaire	5
1 Introduction	1
1.1 Objet de ce travail	1
1.2 Structures et conventions de ce document	5
I CORPUS ET OUTILS	8
2 Corpus	10
2.1 Le <i>Nouveau Corpus d'Amsterdam</i>	10
2.2 BFM : la <i>Base de Français Médiéval</i>	41
2.3 Sources musicales : philologie musicale des XII ^e et XIII ^e s.	48
2.4 Conclusion	63
3 Outils	64
3.1 TXM	64
3.2 Comparalem	67
3.3 Le PAM, <i>Programme d'Analyse Métrique</i>	91
3.4 Conclusion	113
II DÉPOUILLEMENT DES ALTERNANCES GRAPHIQUES	115
4 Noms communs	117
4.1 Introduction	117
4.2 Présentation des résultats	123
4.3 Généralisations	131
4.4 Conclusion	132
5 Adjectifs au féminin singulier	133
5.1 Introduction	133
5.2 Présentation des résultats	136
5.3 Des observations aux généralisations	167
5.4 Conclusion	178
6 Prépositions	180
6.1 Introduction	180
6.2 Présentation des résultats	185
6.3 Des observations aux généralisations	233
6.4 Conclusion	245
7 Adverbes	247
7.1 Introduction	247
7.2 Présentation des résultats	253
7.3 Des observations aux généralisations	288
7.4 Conclusion	298
8 Articles	300
8.1 Introduction	300
8.2 Présentation des résultats	302

8.3	Des observations aux généralisations	323
8.4	Conclusion	325
9	Pronoms personnels	327
9.1	Introduction	327
9.2	Présentation des résultats	333
9.3	Des observations aux généralisations	381
9.4	Conclusion	385
10	Conjonctions : <i>Que, -que</i> et <i>come</i>	388
10.1	Introduction	388
10.2	Présentation des résultats	398
10.3	Des observations aux généralisations	435
10.4	Conclusion	439
11	Chutes des schwas prétoniques	442
11.1	Introduction	442
11.2	Développement des chutes de schwa en hiatus əV	450
11.3	Chute de schwa internes hors hiatus	461
11.4	Discussion	475
11.5	Conclusion	479
III	ANALYSE	480
12	Généralisations sur les données graphiques	482
12.1	Synthèse des taux de défaut	482
12.2	Ce qui tient de l'élision, et ce qui n'en tient pas	488
12.3	Généralisations morphosyntaxiques et phonologiques	496
12.4	Conclusion	507
13	L'élision en métrique et musication	509
13.1	Introduction	509
13.2	L'élision métrique	520
13.3	Naturalité du domaine de l'élision métrique, parophonologie et conflit	529
13.4	Conflit de domaine	534
13.5	Conclusion	536
14	Analyse phonologique	538
14.1	Introduction	538
14.2	Condition phonologique : le flottement	541
14.3	Domaines de l'élision	560
14.4	Conclusion	575
15	Conclusion	577
15.1	Résumé des résultats et analyses	577
15.2	Apports	578
15.3	Ouvertures pour parfaire ce travail	580
ANNEXES		582
A	Abréviations, symboles, notations formelles	583
A.1	Abréviations et symboles	583
A.2	Notations	586

B	Liste des généralisations, règles, principes, etc.	587
C	Corpus informatiques : liste des textes	596
C.1	NCA : liste des textes	596
C.2	BFM : liste des textes cités	608
D	Corpus manuscrits : liste des témoins	609
D.1	Manuscrits des trouvères	609
D.2	Manuscrits d' <i>Aucassin et Nicolette</i> et du <i>Gieus de Robin et de Marion</i>	609
E	Figures rejetées en annexes	610
	Références	636
	Table des matières	652

1.1 Objet de ce travail

Cette thèse s'intéresse à l'origine de l'élision en français. Dans ce travail, nous définissons l'élision comme l'absence d'une voyelle finale atone devant initiale vocalique. Ce processus, qui est toujours actif en français moderne (fr.mod.), était déjà à l'œuvre en ancien français (a.fr.). Il y est attesté par la graphie, par la métrique et par les sources musicales ; il est également largement attesté par les grammairiens du xvi^e s. Ainsi, dans l'incipit de son *Le trehtë de la grammeře françoëze* écrit en alphabet phonétique et cité en épigraphe de ce travail, MEIGRET (1550b, fol. 2v) note <e> pour schwa final devant initiale consonantique (*une pouure confideraçion ; fantazie qe ; une grammeře foët*) mais utilise l'apostrophe devant initiale vocalique (*trop diffiçil' e prę'q'imposibl' en*). Il représente ici la situation par défaut du traitement des finales atones de son époque : elles ne sont supprimées que par l'élision, et se maintiennent en l'absence du contexte prévoicalique qui déclenche l'élision. Mais, comme le montreront nos données, l'élision graphique est beaucoup plus restreinte que l'élision métrique, musicale et que celle des grammairiens. Aussi, la situation par défaut du xvi^e s. que décrit Meigret n'est pas la situation par défaut de l'a.fr., ou du moins pas de toutes les périodes de l'a.fr.

1.1.1 Alternance des voyelles finales atones avec zéro

La situation de l'élision jusqu'à la Renaissance est donc différente de celle du fr.mod., où l'élision ne concerne qu'une série de mots fonctionnels monosyllabiques : *je, me, te, se, ce, de, ne, que, le et la* (TRANEL 2000, p. 59). Cette restriction de l'élision est issue d'une disparition de ses cibles, et ce à double titre. D'une part, en a.fr. et au moins jusqu'au xvi^e s., l'apocope généralisée du schwa final n'est pas encore active, ce qui préserve les schwas et, du même coup, les expose à l'élision devant initiale vocalique. L'élision n'est donc pas nécessairement restreinte aux seuls mots fonctionnels, ainsi qu'en témoignent la métrique, la musication et les grammairiens du xvi^e s. D'autre part, parmi les mots fonctionnels monosyllabiques, les supplétismes qui réduisent les contextes potentiels de l'élision en fr.mod. ne sont pas encore à l'œuvre ; ainsi le déterminant possessif féminin *ma, ta, sa* s'élide (*ma + amie → m'amie, ta + amie → t'amie, sa + amie → s'amie*), au lieu d'être remplacé par la forme masculine (*mon amie, ton amie, son amie*). Aussi, les cibles potentielles de l'élision sont beaucoup plus nombreuses en a.fr. qu'en fr.mod. : toute voyelle finale atone peut, *a priori*, être sujette à l'élision.

Parallèlement, il existe une instabilité marginale de ces finales atones : l'élision n'est pas toujours systématique, et l'on trouve également dans les textes médiévaux des apocopes (absence de voyelle non contextuelle) et des épenthèses (présence d'une voyelle là où il ne devrait pas y en avoir une). Pour ce qui est des irrégularités de l'élision, il faut noter que

certains mots ne sont qu'occasionnellement élidés (*entre els* → *entr els*, *ensemble od* → *ensembl od*), tandis que d'autres ne sont jamais élidés : l'article défini pluriel au cas sujet (CS) *li* ne subit jamais l'élision, alors que l'élision est systématique pour les singuliers *le* (CR) et *la*, et optionnelle pour le cas sujet (CS) singulier *li*.

Quant à l'apocope, elle ne peut être distinguée de l'élision que devant initiale consonantique ; elle est par exemple courante pour certains adverbes, comme « encore » noté *encor* en (1.1). Cet adverbe, formé sur -HÓRA, devrait en principe disposer d'un <e> ; l'absence récurrente de ce <e> devant initiale consonantique indique donc la possibilité de l'apocope.

- (1.1) *n en avoit pas encor tot dit quant le solail resusir vit*
 « elle n'avait pas encore raconté toute l'histoire quand elle vit se soleil se lever »
 (*Vie de Sainte Marguerite*, mrgri, p. 2, trad. pers.)

Enfin, l'épenthèse, comme l'apocope, n'est pas nécessairement dépendante du contexte : un <e> final peut être ajouté sur certains mots et, selon les cas, on peut prêter à ce graphème une valeur phonologique. Les alternances <e>/Ø qui résultent d'une épenthèse peuvent, comme l'élision et l'apocope, se manifester entre différents textes, au sein du même texte, et/ou entre différentes copies du même texte. Ainsi, en (1.2 - 1.4) se trouve le même extrait du *Perceval* de Chrétien de Troyes, dans les ms. F, Q et R, présents dans le NCA (*Nouveau Corpus d'Amsterdam*). Le nombre de syllabes du mot « or » (le métal) varie selon les copies et les occurrences : alors que F dispose bien de deux fois la forme *or*, Q a la forme *or* puis la forme *ore* tandis que R a la forme *ore* puis la forme *or* ; la forme attendue, pour le nom du métal, est *or* (← AÚRUM). Il y a un donc un <e> surnuméraire, à la seconde occurrence en (1.3), et à la première en (1.4).

- | | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>(1.2) vn cheualier armé qui porte
 vne cope d or en sa main
 sa lance teneit et son frein
 et son escu en la senestre
 et la cope d or en la destre
 (<i>Perceval</i>, ms. F, per f, p. 11)</p> | <p>(1.4) [.i.] cheualier armé qui porte
 vne cope d ore en sa main
 sa lance tenoit et son frain
 et son escu en la senestre
 et la cope d or en sa destre
 (<i>Perceval</i>, ms. R, per r, p. 16)</p> |
| <p>(1.3) [.i.] cheualier qui arme porte
 [.i.] cope d or en sa main
 sa lance tenoit et son frain
 et son escu a la senestre
 et la coupe d ore en la destre
 (<i>Perceval</i>, ms. Q, per q, p. 15-16)</p> | <p>(1.5) « un chevalier armé qui emporte
 une coupe d'or dans sa main.
 Il tenait sa lance, ses rênes
 et son écu de la main gauche
 et la coupe d'or de la main droite. »
 (Trad. POIRION in CHRÉTIEN DE TROYES 1994)</p> |

Puisque le mot « or » n'est pas censé, au vu de son étymologie et des principes de la phonétique historique, disposer d'un <e> (AÚRUM évolue en *or*), ce <e> graphique n'est pas de droit. Au-delà de la possibilité du *lapsus calami*, c'est-à-dire de la faute lors de l'écriture, ce <e> peut représenter soit un détachement de la forme graphique vis-à-vis de la représentation sonore du mot, soit une véritable épenthèse vocalique.

1.1.1.1 Distinguer l'élision de l'apocope et de l'épenthèse

Ce travail ne s'intéresse qu'à l'élision, et vise donc à exclure de ses données les cas d'apocope et d'épenthèse. De manière à exclure les cas d'épenthèse, nous avons procédé à la vérification étymologique de tous les lemmes considérés, pour nous assurer que ceux-ci correspondent bien aux critères qui, en phonétique historique, prédisent l'existence d'une finale atone (sec. 1.1.1.2). Pour distinguer l'élision de l'apocope, nous avons systématiquement comparé les taux d'absence du graphème vocalique final devant initiale vocalique et devant initiale consonantique. Ainsi, lorsque nous trouvons une différence statistiquement significative entre les absences de <e>¹ devant initiale vocalique et celles devant initiale consonantique, et que c'est le contexte prévocalique qui est associé aux plus forts taux d'absence de <e>, nous concluons que le processus à l'œuvre est celui de l'élision; si c'est l'inverse ou si la différence n'est pas statistiquement significative, nous ne concluons pas à l'élision.

1: Ou autre graphème vocalique dans le cas des monosyllabes fonctionnels (*la, li*, etc.).

De fait, l'élision, en tant que processus phonologique synchronique, ne concerne pas que les lemmes dont l'étymon doit donner une voyelle finale atone selon la phonétique historique. N'importe quelle voyelle finale épenthétique ou analogique peut être sujette à élision si, dans la représentation lexicale du locuteur, la voyelle est bien présente. Mais, pour filtrer efficacement les données en fonction d'un tel critère, il faudrait établir, pour chaque lemme doté d'une alternance <e>/Ø, si celle-ci est lexicalisée ou non. Nous ne sommes tout simplement pas en mesure de le faire, et encore moins à grande échelle. En revanche, nous pouvons contrôler l'étymologie des lemmes et nous assurer que leur voyelle finale atone est de droit (c'est-à-dire est attendue du point de vue de la phonétique historique). Aussi, à une exception près,² tous les lemmes sur lesquels nous travaillons correspondent à ceux pour lesquels la phonétique historique prédit une finale atone. Les critères retenus, dont la plupart sont communément acceptés, font l'objet de la section suivante.

2: L'exception à ce principe est le mot *come*, dont la spécificité est exposée au chap. 10.

1.1.1.2 Origine des finales atones régulières de l'a.fr.

Une voyelle finale atone de l'a.fr. peut procéder, sauf épenthèse et analogie, d'un étymon latin se terminant en -A (1.6 a), d'un étymon qui a évolué pour comporter en position finale un groupe triconsonantique (1.6 b) ou biconsonantique « non latin » (1.6 c), un groupe *obstruante + liquide* (1.6 d) ou une affriquée /tʃ/ ou /dʒ/ (1.6 e; SÉGÉRAL et SCHEER 2020, GGHF, p. 323-325, 336-345). Une finale atone peut également provenir, dans le cas des mots lexicaux, d'un étymon monosyllabique (1.6 f) ou qui est devenu monosyllabique par aphérèse (1.6 g).

(1.6) Origine des voyelles finales atones de l'a.fr.

- | | | |
|-----|----------|----------------------------------------------------|
| (a) | -A | ROSA → <i>rose</i> |
| (b) | CCC | HÓSPĪTEM → <i>oste</i> ; GRÁCĪLEM → <i>graisle</i> |
| (c) | CC -lat. | CÓMĪTEM → <i>conte</i> ; TĒPĪDU → <i>tiede</i> |
| (d) | OL | PÁTREM → <i>pere</i> ; NÓSTRUM → <i>nostr</i> |
| (e) | /tʃ, dʒ/ | frk. <i>rīki</i> → <i>riche</i> |
| (f) | Mono-σ | ET → <i>e, et</i> ; DE → <i>de</i> |
| (g) | Aph. | ILLI → <i>li</i> (ARTDÉF) |

Dans les autres cas, sauf idiosyncrasie rare, la voyelle finale latine n'a pas de descendance en a.fr. : après une consonne simple (1.7 a), une géminée (1.7 b) ou une voyelle, l'a.fr. n'a pas de finale atone. C'est également le cas après un groupe biconsonantique hétérosyllabique licite en latin, c'est-à-dire un groupe homorganique commençant par une nasale (1.7 c), un groupe rC (1.7 d), lC (1.7 e), sC (1.7 f), jC (1.7 g) ou formé d'occlusives sourdes (1.7 h).

(1.7) Origine des absences de voyelles finales atones en a.fr.

(a)	C	NÓVUM → <i>neuf</i> ; FÍDEM → <i>foi</i>
(b)	C:	DRÁPPUM → <i>drap</i> ; SÍCCUM → <i>sec</i>
(c)	NC	CÁMPUM → <i>champ</i> ; MÚNDUM → <i>mont</i>
(d)	rC	CÓRPUS → <i>cors</i> ; CÓRNU → <i>corn</i>
(e)	lC	SÁLTUM → <i>salt</i> ; SAUT → <i>saut</i> ; FÁLSU → <i>fals, faus</i>
(f)	sC	CRÍSPUM → <i>cresp</i> ; HÓSTEM → <i>ost</i>
(g)	jC	VÁRĪUM → <i>vair</i> ; PÁCE → <i>pais</i>
(h)		SÉPTE → <i>set</i> ; RÉPŪTU → <i>ret</i>

La distinction entre groupes consonantiques latins et non latins permet de remplacer la distinction entre proparoxytons à syncope précoce et proparoxytons à syncope tardive des grammaires traditionnelles. En effet, SAMPSON (1980) a montré que la présence ou l'absence d'un schwa final dans le descendant d'un paroxyton latin était prédictible à partir de l'identité des consonnes que la syncope fait se rencontrer. Lorsque ces consonnes forment un groupe connu du latin, celui-ci est licite en position finale et permet la chute de la finale. En revanche, lorsque le groupe est inconnu du latin, il n'est pas licite en position finale et déclenche l'apparition d'un schwa qui vient remplacer la voyelle finale atone au moment de sa chute. Ainsi, tout groupe consonantique n'appartenant pas à la liste en (1.7 c - h) est illicite en position finale et s'oppose à l'absence de voyelle après lui.

Ce sont ces critères qui nous permettent de définir si un mot, en a.fr., doit *a priori* disposer d'une voyelle finale atone. Lorsqu'une telle voyelle est absente, nous parlons de défaut de <e>, ou de défaut de <V> pour les mots fonctionnels qui ne se terminent pas en schwa.

1.1.2 Projet de cette thèse

Cette thèse est donc dédiée à l'élision en ancien français. En effet, alors qu'on peut constater des irrégularités à la lecture de textes médiévaux, celles-ci ne sont guère prises en compte par les grammaires. De fait, l'élision est un sujet absent d'une partie des grammaires et, parmi les grammaires qui relèvent le phénomène, WALKER (1981, p. 51), MATTE (1982, p. 153), LA CHAUSSÉE (1989 [1974], p. 210), CLÉDAT (1917, p. 67) et PRICE (1985 [1971], p. 76) présentent le phénomène comme stable et invariant, mais sans en donner les conditions précises. POPE (1966 [1934], p. 216) note que l'élision est déjà majoritaire pour *ke* et *se* dans la *Chanson de Roland* mais que la conjonction *ne* se maintient en hiatus au XII^e s. Elle note également que le mot élidé doit entretenir une relation étroite avec le mot suivant. Seul NYROP (1899, p. 229-230) relève la variation du phénomène (« La règle de l'élision souffrait beaucoup d'exceptions ») et

note que la question des contextes syntaxiques et sémantiques précis qui la conditionnent reste « fort embrouillée ».

Si la question était « fort embrouillée » à l'époque de Nyrop, elle ne l'est pas moins aujourd'hui. Cette confusion se fonde notamment sur le fait que, par élision, les auteurs peuvent désigner deux phénomènes : la fusion d'un mot fonctionnel avec le mot suivant (comme en fr.mod.), ou bien la non réalisation de toute voyelle finale atone devant initiale vocalique (c'est le cas en métrique). Nous avons donc voulu éclaircir un peu cette question : cette thèse vise à établir quelles sont les cibles et les conditions de l'élision en a.fr., et selon quelles modalités il peut y avoir variation. Pour ce faire, et face au relatif mutisme des grammairiens, nous avons adopté une approche de linguistique informatique. Nos données proviennent principalement de deux corpus informatiques, le *Nouveau Corpus d'Amsterdam* (NCA; STEIN, KUNSTMANN et GLEßGEN 2006) et la *Base de Français Médiéval* (BFM; GUILLOT-BARBANCE, HEIDEN et LAVRENTIEV 2017). Nous avons développé deux programmes informatiques pour exploiter ces corpus. Le premier, Comparalem, permet l'identification des lemmes et des occurrences sujettes à l'élision graphique. Le second, le PAM (Programme d'Analyse Métrique; POGGIO et PREMAT 2019; PREMAT et POGGIO [à paraître]), permet un repérage des vers irréguliers et donc une détection des élisions métriques qui ne seraient pas transcrites dans la graphie.

L'essentiel de ce travail repose sur l'utilisation de Comparalem avec les données du NCA. L'objectif, ici, sans viser à l'exhaustivité et en tenant compte des limites imposées par le corpus et son annotation, est d'obtenir une vision d'ensemble de la situation de l'élision graphique. Le second programme, ainsi que des relevés et analyses non informatiques, nous permettent de dépasser l'élision graphique pour accéder à l'élision que les scribes ne notent pas.

1.2 Structures et conventions de ce document

1.2.1 Structure

Ce document est divisé en trois parties. La première détaille la composition des corpus utilisés et les outils informatiques utilisés pour les exploiter. La deuxième détaille la distribution des défauts de graphème vocalique final dans le corpus du NCA, avec un chapitre par partie du discours étudiée. Dans chaque chapitre de cette deuxième partie, nous détaillons l'impact du contexte subséquent sur les défauts de <e>, ce qui permet d'établir s'il s'agit d'une élision ou d'une variation non contextuelle. Nous détaillons également des éléments relatifs à la distribution en fonctions des lemmes, et étudions la distribution diachronique et la distribution diatopique. Ce déploiement des données représente, de fait, un volume textuel considérable ; il représente le fondement empirique de notre analyse de l'élision graphique.

La troisième partie produit des généralisations sur les élisions graphiques, les compare avec les élisions de la métrique, de la musique vocale et celles attestées au xvi^e s. chez les grammairiens. Cette partie aboutit à la conclusion qu'il faut supposer, à partir du xiii^e s., une élision plus

large que ne l'est l'élosion graphique. Le dernier chapitre propose une analyse phonologique, diachronique et synchronique, des conditions qui président à l'élosion.

Le sommaire précédant cette introduction est complété, en fin de document, par une table des matières exhaustive. Cette table des matières est précédée d'une liste des figures et tables. Différents éléments sont reportés en annexes : la liste intégrale des abréviations, sigles et conventions, la liste des textes du NCA, la liste des textes de la BFM cités, une liste synthétique des généralisations, règles et processus notés précédemment, ainsi que des figures qui ont été rejetées hors du corps du texte.

1.2.2 Conventions

Les textes issus des corpus numériques sont cités par leur identifiant, en caractères à chasse fixe. Dans la marge, un commentaire non numéroté reproduit des informations concernant ce texte, sauf lorsqu'il a été cité peu de temps auparavant. Les textes qui reviennent le plus souvent dans notre analyse, comme la *Chanson de Roland*, peuvent être cités par leur titre abrégé plutôt que par l'identifiant du texte. Ainsi, *Roland* correspond à *rolandox* dans le NCA ou *roland* dans la BFM. Sauf précision contraire, lorsqu'un texte cité par un titre en italiques (p.ex. *Roland*) existe dans les deux corpus, c'est toujours de l'édition du NCA qu'il est question.

Comme illustré pour *rolandox* et *roland*, la mention de l'identifiant du texte s'accompagne de la note marginale de description. Lorsque le corpus d'origine d'un texte n'est pas précisé, c'est qu'il provient du NCA. Hormis *Roland*,³ les textes versifiés du NCA sont cités sans numéro de vers (absents dans le corpus), et en fonction du numéro de page affiché par le logiciel TXM.

Les marges contiennent également les notes, en numérotation courante par chapitre. Les figures et tables peuvent déborder du corps de texte pour s'étendre dans la marge. Par ailleurs, dans la mesure où ce document contient beaucoup de figures, et où celles-ci peuvent être de grande taille, nous avons tenté de faciliter tant que possible la lecture en les plaçant sur la même page que leur mention, ou du moins à proximité. Pour ce faire, nous tolérons des blancs dans certaines pages pour ne pas trop éloigner une figure de son commentaire.

Lors de la première citation d'une référence bibliographique, ou lorsque la précédente citation d'une référence est éloignée, cette citation est développée dans la marge avec son titre, comme illustré ici pour l'ouvrage de SHORT (2007). Ces développements marginaux ne se substituent pas à la bibliographie finale, et sont uniquement conçus comme une aide à la lecture. Les citations de la *Grande Grammaire Historique du Français* (GGHF; MARCHELLO-NIZIA, COMBETTES et al. 2020) représentent un cas particulier. Nous citons ce document en fonction des parties ou chapitres de la grammaire et non en fonction des sections ou sous-sections. Ainsi, la citation de SÉGÉRAL et SCHEER (2020, GGHF, p. 323-325, 336-345) *supra* fait référence, par ses numéros de page, au chapitre 23 (« Voyelles en syllabe ouverte (CV) : 2. atones », signé par Philippe Ségéral) de la partie 3 (« Phonétique historique », cosignée par Philippe Ségéral et Tobias Scheer). Notre formule de citation réfère donc directement à la partie et non au chapitre, et c'est uniquement la pagination qui indique le

rolandox : ANON. (1100*). *La Chanson de Roland*. Ms. O; Dees : Angleterre; compo. N.-O., ca 1100; ms. a.n., ca 1137; CR1 (SEGRE 1971); DEAF : Ro1S.

roland : ANON. (ca 1100). *Chanson de Roland*. Compo. ca 1100, ms. ca 1120, dial. norm. (MOIGNET 1972a).

3: Pour *Roland*, les numéros de vers sont ceux de l'édition de la BFM, y compris lorsque le texte cité est celui du NCA; nous avons établi la correspondance dans ce cas.

SHORT (2007). *Manual of Anglo-Norman*.

SÉGÉRAL et al. (2020). « Phonétique Historique ».

chapitre voire la section. Ce choix permet de gérer la multiplicité des auteurs de cette grammaire sans multiplier jusqu'à l'absurde les entrées bibliographiques, ni toutes les réduire à une seule entrée générale. Une exception est également faite lorsqu'un auteur a apporté une contribution à un ouvrage qui ne peut correspondre à la rédaction d'un chapitre (p.ex. lorsque le traducteur d'un texte n'est pas l'éditeur du texte, ou lorsque l'éditeur n'est pas l'entrée de la notice, qui est au nom de l'auteur). Dans ce cas, des mentions du type « trad. POIRION *in* CHRÉTIEN DE TROYES 1994 » indiquent la traduction effectuée par POIRION dans le cadre de l'édition des œuvres complètes de Chrétien de Troyes en Pléiade. Les autres citations bibliographiques suivent les normes habituelles.

Les renvois vers les figures et tables sont systématiquement introduits par *figure* ou *fig.* et *table*; un renvoi non précédé d'une telle mention renvoie aux exemples, règles et autres éléments à numérotation courante. Par exemple, (1.7) renvoie à la liste des contextes latins ou gallo-romans n'autorisant pas le maintien d'une voyelle finale atone en a.fr. Ce dernier type de renvoi est placé entre parenthèses, sauf lorsque le renvoi se situe déjà dans des parenthèses.

Une liste complète des abréviations utilisées se trouve en annexes, p. 583 et suivantes. Nous ne reproduisons dans la section suivante que les abréviations bibliographiques; les autres abréviations, à l'exception des abréviations usuelles et des abréviations de noms de langues et de dialectes, sont développées au moins lors de leur première mention en plus de texte.

1.2.2.1 Abréviations bibliographiques

- AND** *Anglo-Norman Dictionary (AND² Online Edition)* (2022).
- DEAF** *Dictionnaire Étymologique de l'Ancien Français (DEAF)* (1974-). La version citée est toujours celle du DEAFél.
- DEAFél** *Dictionnaire Étymologique de l'Ancien Français, version électronique (DEAFél)* (2010-). Cf. DEAF, DEAFpre.
- DEAFpré** Sous ensemble du DEAFél, articles en version préliminaire.
- DECT** *Dictionnaire Électronique de Chrétien de Troyes (DÉCT)* (2015).
- DÉRom** *Dictionnaire Étymologique Roman (DÉRom)* (2022).
- DMF** *Dictionnaire du Moyen Français (DMF)* (2020).
- FEW** *Französisches Etymologisches Wörterbuch (FEW) informatisé* (WARTBURG 2022 [1922-1967]).
- GD** *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle* (GODEFROY 1891-1902).
- GDc** *Complément au Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle* (GODEFROY 1895-1902).
- GGHF** *Grande Grammaire Historique du Français* (MARCHELLO-NIZIA, COMBETTES et al. 2020).
- MGH** *Monumenta Germaniae Historica*, coll. citée par ouvrages dans la bibliographie, cf. BORETIUS et KRAVSE (1887).
- REW** *Romanisches Etymologisches Wörterbuch* (MEYER-LÜBKE 1911-1920).
- TL** *Altfranzösisches Wörterbuch*, version électronique (TOBLER et LOMMATZSCH [2002] 1925-1976).

I
CORPUS ET OUTILS

DANS cette partie, nous détaillons les corpus sur lesquels nous travaillons (chap. 2) ainsi que les outils que nous utilisons pour interroger ces corpus (chap. 3). Nous avons travaillé sur un corpus numérique principal, le *Nouveau Corpus d'Amsterdam* (NCA, sec. 2.1), ainsi que secondairement sur un autre corpus numérique, la *Base de Français Médiéval* (BFM, sec. 2.2) et, de manière plus ponctuelle, sur des sources musicales (sec. 2.3). Le premier corpus numérique est exploité avec TXM (sec. 3.1) et le programme Comparalem (sec. 3.2), tandis que le second est exploité avec le PAM, Programme d'Analyse Métrique (3.3).

L'usage qui est fait de ces différents corpus et programmes dans cette thèse est variable ; l'essentiel du volume textuel de ce document repose sur l'analyse du NCA avec Comparalem. Les résultats de l'exploration de la BFM par le PAM ne sont évoqués que rapidement au chap. 13, dans la mesure où ils confirment la règle et ne révèlent pas de tendance diachronique pertinente. Mais, dans la mesure où ces résultats apparaissent dans le texte de cette thèse, il a semblé nécessaire de donner toutes les informations techniques pertinentes pour ce corpus et ce programme. De même, la musication n'est évoquée que dans ce même chapitre 13, principalement à partir d'exemples commentés. Mais ces exemples ne peuvent être compris que si un minimum d'éléments de philologie musicale a été exposé, aussi sommes nous contraints de développer la composition du corpus utilisé, qui fait appel à des connaissances musicales et historiques précises et parfois contre-intuitives pour le musicien moderne.

Cette thèse se fonde sur l'analyse d'un corpus principal, le *Nouveau Corpus d'Amsterdam* (NCA), un recours ponctuel à un second corpus, celui de la *Base de Français Médiéval* (BFM) ainsi qu'un appel marginal à des sources qui se trouvent hors de ces corpus informatiques, notamment pour ce qui est des sources musicales. Chacun de ces trois ensembles de textes correspond à un usage différent. Nous utilisons le NCA pour les analyses grapho-phonologiques automatisées, grâce à la richesse de son annotation. Nous utilisons la BFM pour documenter la situation du très ancien français et dans les sources versifiées, la BFM disposant d'un corpus plus étendu que le NCA et d'une normalisation des formes graphiques qui permet l'utilisation d'un programme d'analyse métrique. Enfin, les sources musicales servent de fondement à une analyse du traitement des voyelles finales atones dans la musication, c'est-à-dire dans l'association entre syllabes et notes.

Dans ce chapitre, nous décrivons ces corpus, en précisant les propriétés utilisées pour notre analyse, les limites de chaque corpus ainsi que les éléments philologiques nécessaires à la bonne compréhension des sources et éditions. Nous décrivons d'abord le *Nouveau Corpus d'Amsterdam* (sec. 2.1), avant de décrire plus rapidement la *Base de Français Médiéval* (sec. 2.2) et l'ensemble des sources musicales utilisées (sec. 2.3).

2.1 Le *Nouveau Corpus d'Amsterdam*

Le principal corpus sur lequel repose cette thèse est le *Nouveau Corpus d'Amsterdam* (STEIN, KUNSTMANN et GLEßGEN 2006), conventionnellement abrégé NCA. Ce corpus est issu des données de l'*Atlas des formes linguistiques des textes littéraires de l'ancien français* de DEES (1987), mais il ne représente pas exactement le corpus de cet *Atlas* ; de nombreux textes du NCA n'apparaissent pas dans l'atlas, un texte de l'atlas manque au NCA, et l'annotation et la structure des textes ont été repris par l'équipe du NCA. Au sein du NCA, chaque texte dispose de métadonnées (identification, localisation, datation, genre, etc.) et, au sein de chaque texte, chaque mot est annoté en lemmes et en parties du discours. Le corpus est téléchargeable sur demande auprès d'Achim Stein, après signature d'une licence d'utilisation. Il peut être importé dans TXM, qui offre alors des possibilités d'exploitation textométrique standardisées.

Le corpus intégral contient 296 textes, pour un peu plus de 3 180 000 mots. La notion de *texte*, au sein du NCA, doit être précisée : *texte* ne renvoie pas, ici, à une *œuvre*, mais à un texte matériel : différents *textes* peuvent correspondre à autant de versions de la même *œuvre*. Par ailleurs, le NCA ne comprend pas que des textes intégraux, mais contient aussi un certain nombre d'extraits, de longueur variable. Ainsi, par exemple, le roman de Chrétien de Troyes *Yvain, ou Le Chevalier au Lion* est présent en sept versions dans le NCA, qui sont toutes des extraits (vers 1 à 1000 dans toutes les versions sauf une qui s'arrête au v. 993). Chaque

STEIN et al. (2006). *Nouveau Corpus d'Amsterdam*. *Corpus informatique de textes littéraires d'ancien français* (ca 1150-1350).

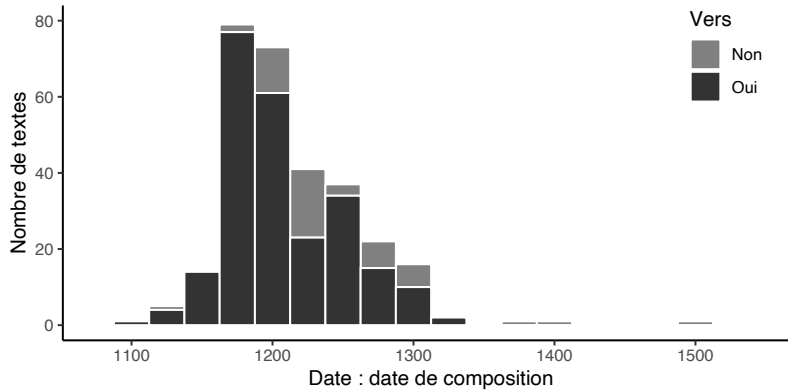


FIGURE 2.1 – Répartition des textes du NCA selon la datation de Dees ou, à défaut, la datation de composition.

version correspond à une copie différente. Il y a donc 7 *textes* (un fichier informatique par version : yva, yvf, yvg, yvh, yvp, yvs et yvv) pour un extrait d'une seule *œuvre littéraire* (*Yvain, ou Le Chevalier au Lion*). Dans ce cas, la dernière lettre de l'identifiant du texte dans le NCA renvoie au sigle du ms. dans la tradition philologique (yva : ms. siglé A par la tradition). Par ailleurs, en quelques cas, plusieurs textes représentent différents extraits de la même *œuvre* : ainsi, mora1 et mora2 représentent la première et la seconde moitié du *Poème moral, Traité de la vie chrétienne* ; cela s'explique par le fait que, en réalité, ces deux extraits d'une même œuvre ne proviennent pas du même ms. Enfin, dans le cas d'*Aucassin et Nicolette*, prosimètre du début du XIII^e s. dont un seul témoin existe, l'œuvre est divisée en deux textes : les sections impaires, versifiées et chantées, forment le texte aucchants, tandis que les sections paires, en prose, forment le texte auc. Au contraire, plusieurs *textes* du NCA rassemblent des extraits provenant de deux œuvres différentes (chret1 et chret2 regroupent chacun les 1000 premiers vers d'*Yvain ou le Chevalier au lion* et les 2000 premiers vers du *Perceval* de Chrétien de Troyes, chaque texte du NCA représentant un manuscrit contenant les deux œuvres). Les 296 textes du NCA ne représentent donc pas 296 œuvres littéraires, mais 296 copies, intégrales ou seulement sous la forme d'un extrait, ou parfois sous la forme de plusieurs extraits de plusieurs œuvres du même auteur contenues dans le même témoin.

Comme représenté en fig. 2.1, le corpus est centré sur des textes composés à la fin du XII^e ou au XIII^e s., la moyenne des dates des textes s'établissant en 1211 ; 242 textes sur 296 sont versifiés (soit 88%). L'absence presque totale de textes non versifiés avant 1200 n'est pas un effet du corpus, mais représente l'apparition puis le développement de l'écriture en prose autour de 1200, dans un paysage toujours dominé par l'écriture versifiée (ZUMTHOR 1972, p. 59-61).

ZUMTHOR (1972). *Essai de poésie médiévale*.

Ainsi que le représente la fig. 2.2, le corpus est dominé par le Nord-Ouest du domaine, avec 41 textes anglo-normands, 47 textes provenant de la Somme et du Pas-de-Calais, et 28 textes normands ; sur cette figure, les régions sont celles de Dees, supplées par la localisation philologique de composition ou, à défaut, de copie, lorsque la localisation de Dees est man-

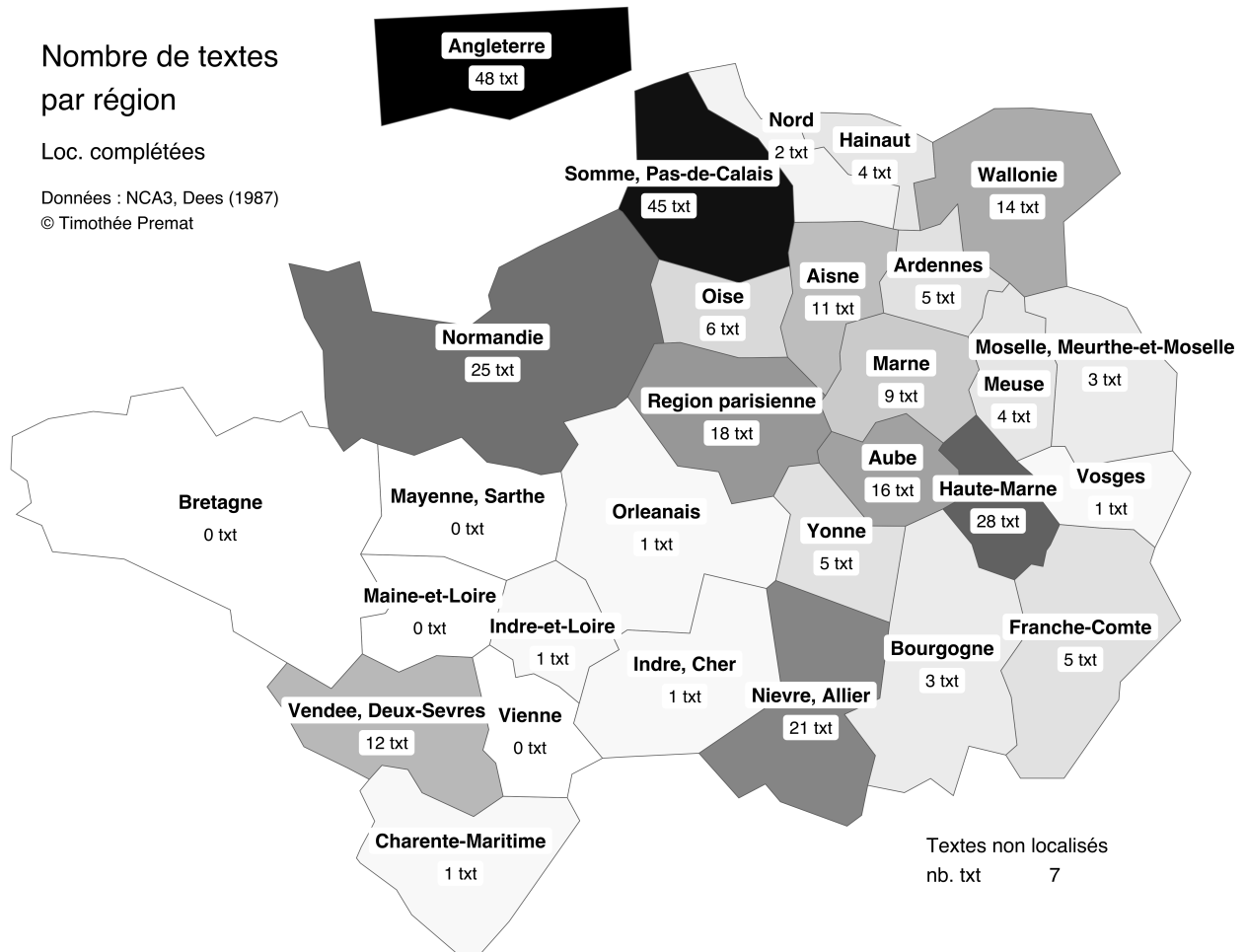


FIGURE 2.2 – Répartition des textes du NCA selon la localisation de composition.

quante (cf. *infra* pour les différents types de localisations). Au contraire, le quart Sud-Ouest du domaine souffre d'une sous-représentation chronique, avec seulement 10 textes localisés en Vendée et Deux-Sèvres, 2 textes en Indre-et-Loire, 1 texte en région Indre, Cher et 1 texte en Charente-Maritime. Le quart Sud-Est et le quart Nord-Est du domaine s'en sortent mieux, d'une part puisqu'aucune région n'y est dépourvue de textes, et d'autre part grâce aux régions Aube et Haute-Marne (23 textes chacune), à la région Nièvre, Allier (20 textes), à la Wallonie (13 textes) et à l'Aisne (11 textes). Il y a donc un certain déséquilibre diatopique, Dees ayant visiblement eu du mal à trouver des documents littéraires de qualité dans le Sud-Ouest. Une nouvelle fois, ce déséquilibre n'est pas qu'un effet du corpus; il représente aussi, ne serait-ce que partiellement, l'état de la production littéraire médiévale, d'abord dominée par le pôle anglo-normand dont la production fut non seulement importante mais surtout précoce (SHORT 2007, p. 32) et par les régions sous l'influence anglo-normande des Plantagénets (en l'espèce, ici, la Normandie), puis par les pôles champenois et picards (LODGE 1997, p. 156) (avec, ici, une sur-représentation de l'Aube au titre de l'ensemble champenois, cf. *infra*).

SHORT (2007). *Manual of Anglo-Norman*.

LODGE (1997). *Le français, histoire d'un dialecte devenu langue*.

Chaque texte dispose de riches métadonnées, au niveau de l'identification du texte, de sa datation et de sa localisation, de son genre ainsi que de la qualité de l'édition utilisée. Dans ce travail, les métadonnées utilisées ont été extraites par nos soins des fichiers XML ; elles comportent ponctuellement des incohérences ou des erreurs de saisie, qui ont été rectifiées. Pour la localisation, deux types de métadonnées sont disponibles : des métadonnées que nous qualifions de dialectométriques¹ établies par Dees (la langue est localisée par un algorithme informatique), et des métadonnées philologiques, qu'une équipe dirigée par Martin-D. Glessgen a ajoutées au corpus.

Du fait d'une certaine hétérogénéité entre ces différentes métadonnées, et de lacunes au sein de certaines d'entre-elles, le corpus sur lequel nous travaillons est légèrement plus restreint que le NCA. Pour exposer les choix que nous avons faits, et pour comprendre l'intérêt et les limites de chacune de ces informations, il est essentiel de reprendre l'histoire de ce corpus et de ses transformations. Dans ce qui suit, nous traitons de la constitution du corpus (sec. 2.1.1), de sa transmission et modernisation informatique (sec. 2.1.2), avant d'aborder l'état des métadonnées vis-à-vis de cette transmission (sec. 2.1.3), pour aboutir à la description et à la distribution des métadonnées retenues pour notre analyse (sec. 2.1.4).

2.1.1 La genèse du NCA : l'Atlas de DEES (1987)

Comme mentionné, le NCA se fonde sur le corpus préparé par Dees pour son *Atlas des formes linguistiques des textes littéraires de l'ancien français* paru en 1987. Cet atlas a paru sous forme imprimée, proposant une série de cartes portant sur un trait graphématique donné, et n'a donné lieu à aucune diffusion informatique. L'Atlas imprimé semble être, dans le projet de Dees, l'objet final, ce qui limite fortement les possibilités de ré-utilisation de cet objet précieux par la communauté après la parution : les chercheurs n'ont accès qu'aux cartes produites. Le *Nouveau Corpus d'Amsterdam* se distingue ici du projet de Dees, et justifie ainsi le qualificatif *nouveau* de son titre : le corpus de Dees (*Corpus d'Amsterdam*), remis en forme aux standards informatiques modernes, est accessible directement, et peut être interrogé par différents outils informatiques, dont la plateforme TXM (HEIDEN, MAGUÉ et PINCEMIN 2010). Or, la transmission des matériaux du corpus de l'Atlas à l'équipe du NCA n'a pas été faite par Dees lui-même, et a eu lieu après le décès du professeur néerlandais, survenu en 2001.

Ainsi, c'est après le décès de Dees que la transmission qui autorise l'existence du NCA a eu lieu : c'est Pieter van Reenen, collaborateur de Dees, qui a transmis à Pierre Kunstmann et Achim Stein des données qu'Anthonij Dees avait préparées pour son *Atlas*. Cet atlas de 1987 s'inscrit dans un projet au long cours qui fut celui de Dees ; il fait suite à plusieurs publications de Dees portant sur l'exploitation de corpus de textes médiévaux, dont la première, son *Étude sur l'évolution des démonstratifs en ancien et en moyen français* (1971), fut saluée par LABOV (1982, p. 36) pour son caractère novateur, encourageant l'emploi de méthodes quantitatives pour la linguistique diachronique. La démarche de Dees ne répond toutefois pas qu'au manque de rapport quantitatif à l'empirie, mais tente également de pallier l'un des problèmes majeurs des études de son temps : « le manque de précisions des données » (KUNSTMANN et STEIN

1: Nous n'avons pas connaissance de l'usage du terme par Dees lui-même, mais ce qualificatif nous semble approprié.

HEIDEN et al. (2010). « TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement ».

LABOV (1982). « Building on empirical foundations ».

2007, p. 10) qui sous-tendent les analyses, c'est-à-dire l'usage de textes mal localisés et mal datés. La simple démarche quantitative ne suffit donc pas, et le corpus de l'*Étude* est également qualitativement contrôlé : il est exclusivement composé de chartes, documents datés et localisés par des indices extra-linguistiques. Dees accomplit ainsi une étude proprement novatrice, tant par son aspect quantitatif que par l'évaluation scrupuleuse des matériaux considérés.

Rejoint par Pieter van Reenen dès 1972, Dees réunit une équipe qui mettra au point ce qui deviendra l'*Atlas des formes et constructions des chartes françaises du 13^e siècle* (1980), renforcé, vis-à-vis du travail précédent, par une informatisation dans le cadre d'un programme de la Vrije Universiteit Amsterdam. Ce sont ainsi 3300 chartes qui ont été numérisées et au moins partiellement annotées. Les chartes présentant un contenu lexical et stylistique par nature borné, l'équipe a ensuite étendu son étude aux textes littéraires, constituant un corpus moins grand en nombre de textes mais pas moins imposant en nombre de mots, les chartes étant par nature des documents brefs. C'est donc sur le modèle de l'*Atlas des chartes* de 1980 que Dees et son équipe publieront l'*Atlas des formes linguistiques des textes littéraires de l'ancien français* de 1987.

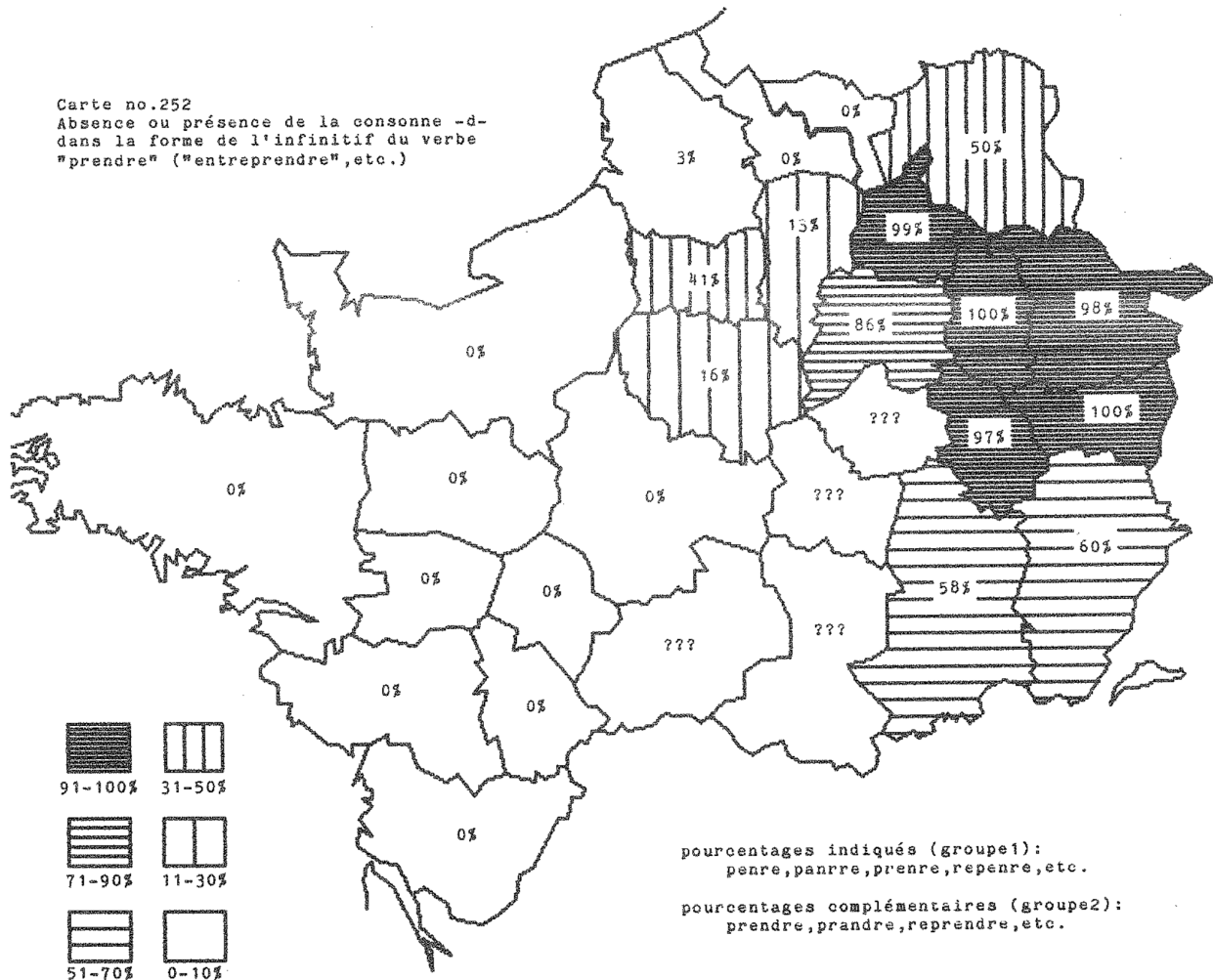
Les deux atlas de Dees reposent sur un principe simple : les textes sont localisés, un trait est identifié (p.ex. épenthèse de /d/ au sein des groupes /nr/), formalisé en tant qu'opposition binaire entre deux variantes ou deux ensembles de variantes (p.ex. présence de /d/ *versus* absence de /d/ au sein des groupes /nr/) et un ensemble de formes répondant aux critères de ce trait est interrogé. Ensuite, une carte est produite, en affichant le pourcentage de l'une des deux variantes pour chaque région. Sous la carte, un tableau reporte pour chaque région le nombre de textes répondant aux critères en question, ainsi que leur division en fonction des variantes, et le score de chaque région. Ce principe est illustré, pour l'épenthèse consonantique dans les groupes /nr/, en fig. 2.3 pour l'*Atlas* de 1980, et en fig. 2.4 pour celui de 1987. Au niveau informatique, ce processus recourt à une base de données dans laquelle, pour chaque texte, chaque forme est dotée d'une étiquette numérique représentant sa partie du discours (cf. 2.1 *infra*).

Si le principe statistique et cartographique est commun aux deux atlas, la conception des métadonnées diffère : alors que le problème de la localisation et de la datation des textes était résolu par le choix du genre des chartes dans l'*Atlas* de 1980, l'*Atlas* de 1987 expose ses auteurs au problème habituel de la philologie médiévale : un texte littéraire peut être composé dans un dialecte donné, ou, au contraire, représenter un mélange de dialectes, et les indices externes de localisation sont rares. De même, la datation pose problème : les œuvres médiévales sont rarement datées dans la source, et quand bien même l'on disposerait d'une date, que daterait-elle ? La première composition de l'œuvre ? Sa copie ultime, vis-à-vis du manuscrit choisi pour l'édition ? Une copie intermédiaire entre les deux ?

Face à ces problèmes, Dees et son équipe ont déployé deux solutions. La première, qui concerne la datation, est relativement simple : les textes doivent correspondre plus ou moins à ce qu'on appelle l'ancien français. En effet, si l'*Atlas* de 1980 comprend quelques cartes diachroniques, ce n'est pas le cas de celui de 1987, qui aborde tous les textes en synchronie.

KUNSTMANN et al. (2007). « Le Nouveau Corpus d'Amsterdam ».

Carte no.252
Absence ou présence de la consonne -d-
dans la forme de l'infinifitif du verbe
"prendre" ("entreprendre", etc.)



Nombre de chartes utilisées

région	total	groupe 1	groupe 2
01 Chte,Chte-Maritime	48	0.00	48.00
02 Vendée,Deux-Sèvres	9	0.00	9.00
03 Vienne	13	0.00	13.00
04 Berry	4	0.00	4.00
05 Orléanais	21	0.00	21.00
06 Indre-et-Loire	8	0.00	8.00
07 Maine-et-Loire	11	0.00	11.00
08 Mayenne,Sarthe	23	0.00	23.00
09 Bretagne	13	0.00	13.00
10 Normandie	19	0.00	19.00
11 Somme,Pas-de-Calais	31	1.00	30.00
12 Oise	39	16.00	23.00
13 Aisne	16	2.00	14.00
14 Nord	168	0.00	168.00
15 Hainaut	33	0.00	33.00
16 Wallonie	26	13.00	13.00
17 Ardennes	38	37.50	0.50
18 Marne	22	19.00	3.00
19 Région parisienne	28	4.50	23.50
20 Yonne	4	1.00	3.00
21 Aube	2	2.00	0.00
22 Haute-Marne	37	36.00	1.00
23 Meuse	30	30.00	0.00
24 Moselle,Mthe-et-Mlle	33	32.50	0.50
25 Vosges	9	9.00	0.00
26 Franche-Comté	26	15.50	10.50
27 Bourgogne	12	7.00	5.00
28 Bourbonnais,Nièvre	1	0.00	1.00
total	724	226.00	498.00

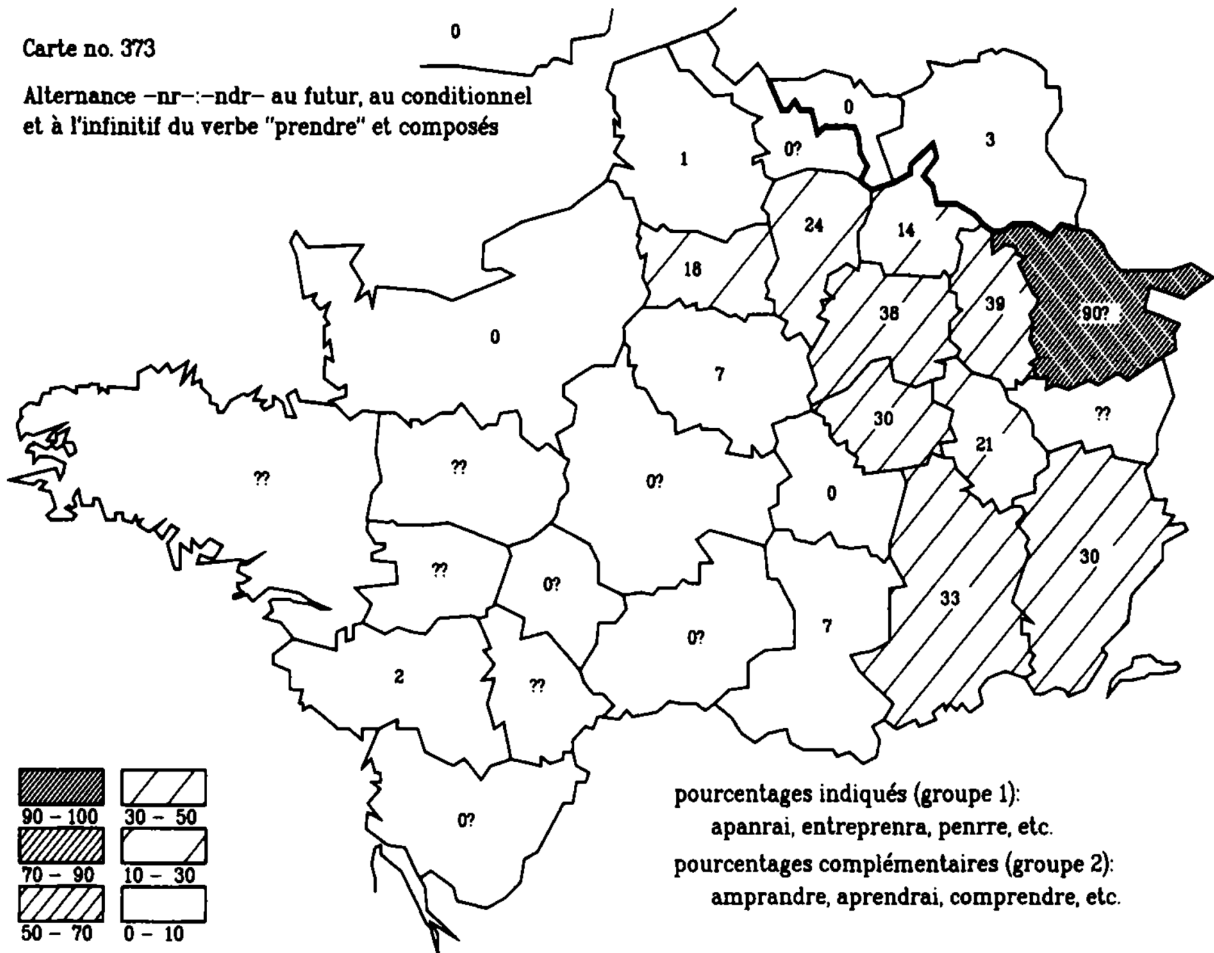
Pourcentages

région	groupe 1	groupe 2
01 Chte,Chte-Maritime	0.00	100.00
02 Vendée,Deux-Sèvres	0.00	100.00
03 Vienne	0.00	100.00
04 Berry	????	????
05 Orléanais	0.00	100.00
06 Indre-et-Loire	0.00	100.00
07 Maine-et-Loire	0.00	100.00
08 Mayenne,Sarthe	0.00	100.00
09 Bretagne	0.00	100.00
10 Normandie	0.00	100.00
11 Somme,Pas-de-Calais	3.23	96.77
12 Oise	41.03	58.97
13 Aisne	12.50	87.50
14 Nord	0.00	100.00
15 Hainaut	0.00	100.00
16 Wallonie	50.00	50.00
17 Ardennes	98.68	1.32
18 Marne	86.36	13.64
19 Région parisienne	16.07	83.93
20 Yonne	????	????
21 Aube	????	????
22 Haute-Marne	97.30	2.70
23 Meuse	100.00	0.00
24 Moselle,Mthe-et-Mlle	98.48	1.52
25 Vosges	100.00	0.00
26 Franche-Comté	59.62	40.38
27 Bourgogne	58.33	41.67
28 Bourbonnais,Nièvre	????	????

FIGURE 2.3 – Carte n° 252 de l'Atlas des formes et constructions des chartes françaises du 13^e siècle (DEES 1980, p. 266).

Carte no. 373

Alternance *-nr-*/*-ndr-* au futur, au conditionnel
et à l'infinitif du verbe "prendre" et composés



Nombre de textes utilisés				Pourcentages			
région	total	groupe 1	groupe 2	région	groupe 1	groupe 2	
01 Charente, Charente-M maritime	1	0.00	1.00	01 Charente, Charente-M maritime	0.00	100.00	
02 Vendée, Deux-Sèvres	5	0.11	4.89	02 Vendée, Deux-Sèvres	2.22	97.78	
03 Vienne	0	—	—	03 Vienne	—	—	
04 Indre, Cher	1	0.00	1.00	04 Indre, Cher	0.00	100.00	
05 Orléanais	1	0.00	1.00	05 Orléanais	0.00	100.00	
06 Indre-et-Loire	1	0.00	1.00	06 Indre-et-Loire	0.00	100.00	
07 Maine-et-Loire	0	—	—	07 Maine-et-Loire	—	—	
08 Mayenne, Sarthe	0	—	—	08 Mayenne, Sarthe	—	—	
09 Bretagne	0	—	—	09 Bretagne	—	—	
10 Normandie	16	0.00	16.00	10 Normandie	0.00	100.00	
11 Somme, Pas-de-Calais	35	0.47	34.53	11 Somme, Pas-de-Calais	1.34	98.66	
12 Oise	5	0.92	4.08	12 Oise	18.49	81.51	
13 Aisne	10	2.36	7.64	13 Aisne	23.62	76.38	
14 Nord	2	0.00	2.00	14 Nord	0.00	100.00	
15 Hainaut	3	0.00	3.00	15 Hainaut	0.00	100.00	
16 Wallonie	11	0.38	10.62	16 Wallonie	3.42	96.58	
17 Ardennes	5	0.68	4.32	17 Ardennes	13.67	86.33	
18 Marne	7	2.64	4.36	18 Marne	37.66	62.34	
19 Région parisienne	14	0.97	13.03	19 Région parisienne	6.96	93.04	
20 Yonne	5	0.00	5.00	20 Yonne	0.00	100.00	
21 Aube	10	3.02	6.98	21 Aube	30.25	69.75	
22 Haute-Marne	19	4.03	14.97	22 Haute-Marne	21.20	78.80	
23 Meuse	3	1.18	1.82	23 Meuse	39.39	60.61	
24 Moselle, Meurthe-et-Moselle	2	1.80	0.20	24 Moselle, Meurthe-et-Moselle	90.00	10.00	
25 Vosges	0	—	—	25 Vosges	—	—	
26 Franche-Comté	4	1.21	2.79	26 Franche-Comté	30.21	69.79	
27 Bourgogne	3	1.00	2.00	27 Bourgogne	33.33	66.67	
28 Nièvre, Allier	20	1.41	18.59	28 Nièvre, Allier	7.05	92.95	
29 Angleterre	8	0.00	8.00	29 Angleterre	0.00	100.00	

FIGURE 2.4 – Carte n° 373 de l'Atlas des formes linguistiques des textes littéraires de l'ancien français (DEES 1987, p. 373).

La seconde solution, qui concerne la localisation, est autrement plus complexe et est tout à fait originale. À partir de l'*Atlas* des chartes, Dees a défini ce qu'on pourrait appeler un *espace dialectométrique*, c'est-à-dire un espace multidimensionnel qui correspond non pas à l'espace géographique mais à l'espace défini par la langue des chartes. Cet espace dialectométrique est ancré dans l'espace géographique en 87 points, qui correspondent à des sous-régions de l'*Atlas* des chartes pour lesquelles l'*Atlas* des chartes dispose de suffisamment de matériel. Chaque dimension de l'espace dialectométrique est formée par un trait linguistique des chartes, auquel est associé son taux pour chaque variante dans chaque sous-région; ces traits sont au nombre de 268. Une fois cet espace dialectométrique défini, chaque texte littéraire est comparé avec les 87 points de l'espace dialectométrique, sur chacune des dimensions de ce dernier : chaque texte littéraire est évalué en termes de proximité avec chaque point pour chaque trait linguistique. Dees en tire un « degré de parenté » DEES (1987, p. xx), c'est-à-dire une valeur numérique qui indique à quel point ce texte est proche de la langue des chartes de cet endroit. La matrice par laquelle il calcule ce degré de parenté comporte donc 87 colonnes et 268 lignes, soit 23 316 valeurs; le degré de parenté fait la synthèse de ces valeurs. Tous les traits n'ayant pas la même valeur quant à la localisation, cet indice de degré de parenté est pondéré par un coefficient dont la valeur est inverse à l'extension géographique du trait dans les chartes : un trait présent sur une large partie du domaine a un poids faible, tandis qu'un trait qui ne serait présent que dans une seule région obtient un poids fort (p.ex. le trait d'insertion de <h> dans les pronoms personnels sujets *ilh*, *ihl* ~ *il(s)* ne concerne que la Wallonie, et sa pondération est donc maximale (DEES 1987, p. xxiv)). Cette pondération produit un « coefficient de dialectalité » (DEES 1987, p. xxi, *et sq.*), qui indique la proximité entre la langue du texte littéraire évalué et les caractéristiques linguistiques des chartes du point disposant du meilleur degré de parenté. Un coefficient de dialectalité de 100%, qui n'est pas attesté dans les textes de l'*Atlas*, indiquerait que la langue du texte correspondrait exactement avec la langue des chartes d'un point donné de l'espace dialectométrique.

Les localisations proposées par Dees doivent donc s'entendre comme étant plus abstraites que les localisations philologiques : il n'est pas ici question de savoir dans quel *scriptorium* le texte aurait été copié, c'est-à-dire dans un point du monde physique, mais bien de savoir de quel point de l'espace dialectométrique (formé par l'intersection entre les 268 traits linguistiques sélectionnés et les 87 points de données sélectionnés) le texte en question se rapproche le plus. Il est ici question d'un double niveau d'abstraction : Dees décide de la localisation en termes de proximité, et de proximité avec un espace dialectométrique, composé de textes d'une nature fort différente. Si cette méthode n'a pas la précision d'éventuels indices extra-linguistiques (p.ex. la mention d'un lieu de copie, la présence attestée d'un auteur au sein d'une cour au moment de la rédaction, associée à une dédicace au patron ou à la patronne de ladite cour, etc.), elle permet en revanche de juger de la langue du texte dans son ensemble. L'origine de l'auteur, le ou les dialectes qu'il utilise pour composer son texte, et l'itinéraire philologique des copies successives sont écrasés en une description homogène de la langue du manuscrit édité. Un texte dont la langue serait trop mélangée, par exemple par un effet de palimpseste entre la langue du premier

rédacteur et la langue du dernier copiste, se signalera par un trop faible coefficient de dialectalité ; c'est le cas de 35 textes, rejetés hors de l'*Atlas* par un coefficient de dialectalité inférieur à 70 (DEES 1987, p. xvii), seuil en-deçà duquel Dees ne considère plus une localisation comme suffisamment fiable. Au contraire, en présence de différentes variantes peu contradictoires, le texte est localisé au mieux, et son coefficient de dialectalité en témoigne de manière objective.

Ce faisant, Dees obtient donc une localisation méthodique, qui parvient en cela à échapper à « l'impressionisme » (DEES 1987, p. xxviii) des localisations philologiques traditionnelles. Si la plupart de ces localisations traditionnelles sont bien confirmées par la méthode de Dees, il n'en demeure pas moins qu'elles reposent sur une méthode instable, changeante de philologue en philologue, et souvent insuffisamment documentée. Le linguiste scrupuleux, s'il n'est pas philologue lui-même, n'a aucun moyen de savoir s'il peut faire confiance à une localisation philologique traditionnelle. Les localisations proposées par Dees, elles, sont méthodiques, et, qui plus est, leur association avec un coefficient de dialectalité donne une idée de la mixité ou de la 'pureté' dialectale de la langue du texte, ce qui est impossible à mesurer sans approche quantitative.

2.1.2 De l'*Atlas* au NCA

L'*Atlas* produit par Dees n'est pas interrogeable : le lecteur n'a accès qu'aux requêtes qu'a effectuées Dees, aux taux de chaque région, à la représentation cartographique de ces taux et à une longue liste de formes non référencées en annexes. C'est pour permettre de nouvelles interrogations du corpus que les matériaux de Dees, transmis sur disquettes par Pieter van Reenen à Pierre Kunstmann et Achim Stein, ont été recyclés en un corpus moderne, encodé en XML et facilement accessible : le NCA, *Nouveau Corpus d'Amsterdam* (STEIN, KUNSTMANN et GLEISGEN 2006). Depuis peu, une interrogation cartographique est disponible en ligne, dans le cadre du projet ADE, *l'Atlas Dees Electronique*, dont la version beta fut disponible en 2022 (SCHEER et al. 2022).

Le recyclage des données de Dees au sein d'un corpus informatique moderne ne se satisfait pas de reproduire à peu près à l'identique les données de l'*Atlas* de Dees dans un format aux normes du début du XXI^e s. Le corpus a en effet été enrichi, et ce dans trois dimensions distinctes : l'annotation des textes, les métadonnées des textes, et le nombre de textes. Concernant l'annotation, les données de Dees se limitent à la forme linguistique directement suivie d'un code numérique correspondant à la partie du discours de la forme, et, en l'absence de balises idoines, le contenu n'est pas localisé dans le texte. Par comparaison, la troisième version du NCA (NCA3) insère chaque mot dans un environnement XML (<w>), et le dote de balises d'annotation qui reprennent l'annotation de Dees (deespos), la développe sous une forme lisible (pos),² la met en concurrence avec une annotation en parties du discours produite par TreeTagger (taggerpos), et la complète par la lemmatisation proposée par TreeTagger (lemma) ainsi que par la ou les sources du ou des lemmes proposés (src) (2.1b).³ Par ailleurs, le contenu est localisé au sein du texte, au moyen de balises phrastiques <s> détournées.⁴ Le développement des étiquettes de Dees s'est au passage enrichi d'un certain nombre de désambiguïisations ; ainsi, l'adverbe *or* (« maintenant »)

SCHEER et al. (2022). *Atlas Dees Électronique* 22.

2: La forme exacte des attributs peut varier selon les traitements qu'a subi le corpus. Dans la version NCA3 pour TXM sur laquelle nous avons travaillé, par exemple, les ':' qui subdivisent les attributs pos sont remplacés par des tirets bas '_', ce qui n'est pas le cas dans la version NCA4 pour TXM.

3: Les balises src ont été retirées en (2.1b) et (2.1c) pour améliorer la lisibilité, leur contenu étant nul.

4: L'unité de découpage <s n="nca:28337_auc:1"> du texte auc correspond ici au premier paragraphe du deuxième chapitre du texte, auc ne contenant, dans le NCA, que les parties en prose du texte, soit les chapitres paires d'*Aucassin et Nicolette*. Selon les recommandations de la TEI-P5, l'élément <s> correspond plus ou moins à une phrase : « <s> (s-unit) contains a sentence-like division of a text » (<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-s.html>). Ce détournement des balises phrastiques n'est, semble-t-il, pas rare dans la philologie numérique médiévale.

et la conjonction *et* disposent de la même étiquette 311, alors que leur reprise sous la forme de l'attribut de pos dans le NCA les distingue, *or* étant annoté ADV et *et* CON: coord. Ce sont visiblement les annotations de TreeTagger qui ont servi à lever ce genre d'ambiguïtés, mais sans écraser les informations de Dees dans le cas où l'étiquette de Dees est plus précise que celle de TreeTagger, comme pour *dient* et *content*, où la combinaison de l'astérisque et de l'étiquette 516 indiquent qu'il s'agit d'un verbe conjugué à la troisième personne du pluriel de l'indicatif présent (VER:pres:3:pl). Enfin, la version 4 du NCA (NCA4), sortie en 2023, ajoute une nouvelle annotation en parties du discours et en lemmes, produite par le programme RRN tagger, spécialement entraîné à partir de données hautement qualitatives issues de la BFM (*BFM Gold Standard data*); l'annotation de TreeTagger est conservée avec le préfixe *tt* (2.1c). L'acuité supérieure de RRN tagger peut s'observer pour *content*, que RRN tagger identifie comme le verbe *conter*, alors que TreeTagger ne sait décider s'il s'agit de *conter* ou de *contendre*.⁵ À chaque fois, les informations du ou des précédents corpus sont donc conservées : le NCA3 reproduit l'annotation de Dees, et le NCA4 conserve l'annotation du NCA3 et, donc, celle de Dees. Enfin, il faut noter que le balisage des vers n'a pas été systématiquement introduit dans le NCA : si certains textes sont bien balisés en vers (p.ex. au moyen de balises phrastiques <s> dans *rolandox*, cf. fig. E.1 en annexes, p. 610), ce n'est pas le cas de tous les textes versifiés. Dans *aucchants*, par exemple, ni environnement ni balise ponctuelle ne viennent signaler les frontières de vers, et l'environnement phrastique <s> sert à délimiter l'ensemble de la laisse.

Concernant les métadonnées des textes, elles se résument dans l'*Atlas* à de brèves notices bibliographiques organisées par régions (DEES 1987, p. 519-533), comportant le coefficient de dialectalité et la sous-région auquel il s'applique. Dans le NCA, chaque texte est précédé d'un préambule plus riche, ayant été complété notamment par Martin-D. Glessgen, Xavier Gouvert et Claire Vachon. Ces informations complètent celles données par Dees, sur le plan de l'identification du texte, de ses caractéristiques formelles, de la qualité du travail éditorial, et de ce que la bibliographie philologique dit à propos de la datation et de la localisation du texte (cf. sous-section 2.1.3 *infra*).

Enfin, en ce qui concerne les textes eux-mêmes, les corpus divergent. D'une part, un texte manque, semble-t-il irrémédiablement, parmi les données de Dees qui ont été récupérées : le texte que l'ADE22 appelle *ren1*.⁶ D'autre part, le NCA contient bien plus de textes que l'*Atlas* : il contient non seulement les 35 textes que Dees a écartés en raison d'un coefficient de dialectalité inférieur à 70, mais il contient également des textes écartés pour d'autres raisons (p.ex. *rolandox*, édition du manuscrit d'Oxford de la *Chanson de Roland*, n'a évidemment pas sa place dans un atlas centré sur la seconde moitié du XII^e s. et sur le XIII^e s.). Enfin, le découpage de certains textes n'est pas identique : le NCA découpe *Aucassin et Nicolette* en deux textes, selon que les parties sont versifiées et chantées (sections impaires : *aucchants*) ou sont en prose (sections paires : *auc*), tandis que l'*Atlas* ne dispose que d'une entrée bibliographique, et *pera* et *yvh* forment deux entrées différentes dans l'*Atlas* comme dans le NCA, mais sont également réunis sous *chret2* dans le NCA.

5: *Contendre* : « combattre, lutter, être en lutte, en contestation » (GD).

6: La version de *ren1* présente dans l'ADE22 a été reconstruite à partir de l'édition utilisée par Dees et des moyens modernes d'annotation ; il ne s'agit donc pas du fichier de Dees, qui n'a, à ce stade, pas été retrouvé.

(2.1) Comparaison des annotations entre les données de DEES (1987) (a), le NCA3 (b) et le NCA4 (c).

(a) or_311 *dient_516 et_311 *content_516

(b) NCA3

```

1 | <s n="nca:28337_auc:1">
2 | <w>
   | <form>or</form>
   | <ana type="pos">ADV</ana>
   | <ana type="deespos">311</ana>
   | <ana type="taggerpos">ADV</ana>
   | <ana type="lemma">or4|or</ana>
   | </w>
3 | <w>
   | <form>dient</form>
   | <ana type="pos">VER:pres:3:pl</ana>
   | <ana type="deespos">516*</ana>
   | <ana type="taggerpos">VER</ana>
   | <ana type="lemma">dire</ana>
   | </w>
4 | <w>
   | <form>et</form>
   | <ana type="pos">CON:coord</ana>
   | <ana type="deespos">331</ana>
   | <ana type="taggerpos">CON:coord</ana>
   | <ana type="lemma">et</ana>
   | </w>
5 | <w>
   | <form>content</form>
   | <ana type="pos">VER:pres:3:pl</ana>
   | <ana type="deespos">516*</ana>
   | <ana type="taggerpos">VER</ana>
   | <ana type="lemma">contendre|conter</ana>
   | </w>

```

(c) NCA4

```

1 | <s n="nca:28337_auc:1">
2 | <w>
   | <form>or</form>
   | <ana type="pos">ADV</ana>
   | <ana type="deespos">311</ana>
   | <ana type="ttpos">ADV</ana>
   | <ana type="tlemma">or4|or</ana>
   | <ana type="rnnpos">ADVgen</ana>
   | <ana type="rnnlemma">or</ana>
   | </w>
3 | <w>
   | <form>dient</form>
   | <ana type="pos">VER:pres:3:pl</ana>
   | <ana type="deespos">516*</ana>
   | <ana type="ttpos">VER</ana>
   | <ana type="tlemma">dire</ana>
   | <ana type="rnnpos">VERc|g</ana>
   | <ana type="rnnlemma">dire</ana>
   | </w>
4 | <w>
   | <form>et</form>
   | <ana type="pos">CON:coord</ana>
   | <ana type="deespos">331</ana>
   | <ana type="ttpos">CON:coord</ana>
   | <ana type="tlemma">et</ana>
   | <ana type="rnnpos">CONcoo</ana>
   | <ana type="rnnlemma">et</ana>
   | </w>
5 | <w>
   | <form>content</form>
   | <ana type="pos">VER:pres:3:pl</ana>
   | <ana type="deespos">516*</ana>
   | <ana type="ttpos">VER</ana>
   | <ana type="tlemma">contendre|conter</ana>
   | <ana type="rnnpos">VERc|g</ana>
   | <ana type="rnnlemma">conter</ana>
   | </w>

```

2.1.3 État des métadonnées

Le NCA est donc un corpus composite, et l'on peut semble-t-il distinguer au moins trois strates de textes : (a) les textes préparés et sélectionnés par Dees pour la réalisation de l'*Atlas* de 1987, (b) les textes préparés par Dees mais non sélectionnés pour la réalisation de l'*Atlas* à cause de leur coefficient de dialectalité trop faible, et (c) les textes préparés par Dees mais exclus pour d'autres raisons. Les textes de la catégorie (a) sont au nombre de 200, s'étendent de la seconde moitié du *xii*^e à la fin du *xiv*^e s., et sont dotés d'un coefficient dialectal supérieur ou égal à 70 ; c'est-à-dire que les calculs de Dees assurent aux textes en question une localisation à la fois robuste et précise dans l'espace dialectométrique des chartes de l'*Atlas* de 1980. Les textes de la catégorie (b) correspondent à des textes de la même période mais dont le coefficient dialectal est inférieur à 70 (leur localisation dialectométrique est peu satisfaisante), et sont au nombre de 35. La composition de la catégorie (c) est plus confuse : si l'exclusion est parfois évidente (cf. rolandox *supra*), elle n'est pas documentée.

Les textes des deux premières catégories sont en principe dotés des métadonnées établies par Dees. Celles-ci sont représentées en table 2.1, selon l'en-tête des variables utilisé dans le NCA, et illustrées sur le texte *abe*.

TABLE 2.1 – Métadonnées des textes des catégories (a) et (b).

Variable	Valeur de l'attribut	Exemple d'attribut
id	Sigle alphanumérique	abe
titreDees	Titre du texte	J. de Meun, Traduction de la première épître de P. Abélard, v. 1-821
editionDees	Référence de l'édition	éd. Ch. Charrier, Paris 1934
manuscritDees	Référence du ms.	Paris, Bibl. Nat., fr. 920
regionDees	Localisation du texte	Région parisienne
codeRegional	Code correspondant à <code>regionDees</code>	54
coefficientRegionalDees	Fiabilité de la localisation	84 (Région parisienne)
dateMoyenneDees	Indication chronologique	1325

Le coefficient dialectal est augmenté, entre parenthèses, de la sous-région définie par Dees. Ces sous-régions sont des subdivisions irrégulières de l'espace dialectométrique des chartes, utilisées pour la localisation des textes littéraires, mais elles n'ont pas été utilisées pour les cartes des deux atlas. Le cas classique est celui d'une sous-région incluse dans une région : le texte *abreja* est localisé, au niveau des régions, en Franche-Comté,⁷ et, au niveau des sous-régions, dans le Doubs.⁸ Mais ces sous-régions ne sont pas toujours équivalentes : pour certains textes, la sous-région n'est qu'une recopie de la région (cf. *abe* en table 2.1), tandis que pour les textes dotés d'une véritable sous-région, certaines sont géographiquement discrètes et peuvent être facilement traitées (le plus souvent : nom de départements ou provinces (Belgique) actuels), tandis que d'autres sont des approximations autour d'un point géographique.⁹ Pas plus que Dees, nous ne ferons usage de ces sous-régions, dans la mesure où elles nous exposent à une trop fine granularité qui rendrait l'ensemble difficilement exploitable dans le cadre de notre projet.

Les textes appartenant à la deuxième catégorie (exclus de l'*Atlas* à cause de leur faible coefficient de dialectalité) et à la troisième catégorie (préparés par Dees mais exclus de l'*Atlas* pour une autre raison que le seul coefficient de dialectalité) disposent tous des variables `id` et `titreDees`, mais les autres variables peuvent être manquantes.¹⁰ Dans la version du NCA3 sur laquelle nous avons travaillé, cela concerne 98 textes. Parmi ces 98 textes, 28 disposent d'une `dateMoyenneDees`, 59 disposent d'un `codeRegional` (on peut donc rétablir leur `regionDees`) et 36 disposent d'un `coefficientRegional`. Ces trois sous-ensembles ne se recoupent que partiellement : si tous les textes disposant d'un `coefficientRegional` disposent évidemment d'un `codeRegional`, l'inverse n'est pas vrai (23 textes disposent d'un `codeRegional` mais n'ont pas de `coefficientRegional`), et le lien entre présence d'une `dateDees` et présence d'un `codeRegional` est imparfait : sur les 28 textes disposant d'une `dateDees`, 1 texte ne dispose pas d'un `codeRegional`, et sur les 59 textes disposant d'un `codeRegional`, 31 n'ont pas de `dateDees`. Cet ensemble est donc fortement hétérogène au niveau des métadonnées de Dees.

En revanche, pour tous les textes, le travail philologique et bibliogra-

7: `regionDees` = Franche-Comté.

8: `coefficientRegionalDees` = 86 (Doubs).

9: Pex. *vergih* : `RegionDees` = Haute-Marne; `coefficientRegionalDees` = 85 (Chaumont en env.). Pour plus de détails sur les subdivisions régionales, cf. ADE22, page « 3. Géographie », §3.3.

10: Auquel cas le fichier XML donne à la variable l'attribut `nil`.

phique entrepris par Glessgen et son équipe (cf. GLEßGEN et GOUVERT 2007; GLEßGEN et VACHON 2013) assigne des attributs à des variables plus ou moins équivalentes : ils créent deux variables pour la datation et la localisation, selon la composition ou selon la copie de la version éditée. Ainsi, pour *abe*, le texte qui nous servait d'exemple en table 2.1, les informations sont reprises en (2.2). Chaque information est sourcée, la source la plus courante étant le DEAF.

(2.2) Métadonnées philologiques pour *abe*

(a) <code>DateComposition</code>	1280ca
(b) <code>DateManuscrit</code>	1395ca
(c) <code>lieuComposition</code>	frc. [= francien]
(d) <code>lieuManuscrit</code>	Paris

Ces informations philologiques, ayant été établies pour le NCA, sont donc bien moins lacunaires ; elles ne manquent que lorsque l'information est impossible à établir en fonction de la bibliographie consultée. La `DateComposition` manque pour trois textes,¹¹ la `DateManuscrit` pour 5 textes.¹² Le `lieuComposition` manque pour 38 textes ; le `lieuManuscrit` pour 28 textes. Ces informations sont en revanche de nature très hétérogène : la plupart des dates sont relativisées par un *circa* ou par une fourchette allant de 1 à 100 ans.¹³ L'hétérogénéité est plus problématique encore concernant les localisations : celles-ci peuvent être dialectales¹⁴ ou géographiques,¹⁵ d'une granularité fine¹⁶ ou grossière¹⁷, d'une fiabilité incertaine,¹⁸ voire se résumer à un marquage dialectal partiel¹⁹ et/ou mixte.²⁰ Ces variables `lieu[...]` ne correspondent donc pas à une localisation, mais à une information seulement *relative* à la localisation, qu'elle soit dialectale ou géographique, qu'elle concerne partiellement ou entièrement la langue du texte, et dont la précision va du quart du domaine oïlique à une ville précise. Cette information semble conçue pour un lecteur humain, et non pour un traitement informatique, dans la mesure où elle est faiblement standardisée.

Ce travail philologique juge aussi de la qualité de l'édition utilisée par Dees : l'équipe de Glessgen a attribué à chaque texte une étiquette bipartite, d'abord selon que l'édition est diplomatique (MS) ou critique (CR), et ensuite selon que l'édition est peu (1), moyennement (2), ou très interventionniste (3). Ainsi, les éditions les plus fidèles au témoin sont étiquetées MS1, tandis que les moins fidèles au témoin sont étiquetées CR3 ; dans la version du NCA sur laquelle nous travaillons, le NCA3, de nombreuses éditions sont qualifiées uniquement en fonction du type d'édition et non de la qualité, à savoir MS ou CR. La composition du corpus en termes d'étiquettes qualitatives est reprise en table 2.2, table qui indique aussi le nombre de textes utilisés pour l'*Atlas* de 1987.

2.1.4 Décomposition du corpus en fonction des métadonnées

L'implication mutuelle entre les attributs précédemment cités est faible : si un seul texte dispose d'une `regionDees` sans `codeRegional`, 39 textes disposent d'un `codeRegional` sans `regionDees`. L'implication entre `dateMoyenneDees` et `regionDees` ou `codeRegional` est également faible : si 160 textes disposent des trois métadonnées, 37 textes ont une

11: arr, oakbook, verite

12: bodo, oakbook, rombriva, romi, romo

13: P.ex. greg : `dateComposition` et `dateManuscrit` = 1215pm1; rome : `dateComposition` = 1188pm13, `dateManuscrit` = 1400pm100.

14: P.ex. guib : `lieuComposition` = frc. [= francien].

15: P.ex. sully2 : `lieuComposition` = La Rochelle.

16: P.ex. sully2 : `lieuComposition` = La Rochelle.

17: P.ex. sept : `lieuComposition` = Sud-Ouest.

18: P.ex. amo : `lieuManuscrit` = pic.?

19: P.ex. sage : `lieuComposition` = traits de l'Ouest.

20: P.ex. myst : `lieuComposition` = traits. occ. et agn.

TABLE 2.2 – NCA : qualité des éditions

qualité	nb. textes	dt. DEES (1987)
MS1	85	65
MS1-2	1	0
MS2	33	22
MS3	7	7
MS	81	42
CR1	26	17
CR2	27	23
CR3	13	10
CR	17	11
NIL	6	3

regionDees et un codeRegional mais pas de dateMoyenneDees, 1 texte a une regionDees et une dateMoyenneDees mais pas de codeRegional, 27 textes ont un codeRegional et une dateMoyenneDees mais pas de regionDees, et un seul texte dispose d'une dateMoyenneDees alors qu'il n'a ni regionDees ni codeRegional. Aucune implication mutuelle liée à la stratigraphie du corpus n'est à rechercher du côté des données philologiques dateComposition, dateManuscrit, lieuComposition et lieuManuscrit, leur défaut ne provenant que d'une impossibilité à établir l'information à partir de la bibliographie philologique disponible. De même, il n'y a pas lieu de chercher une implication mutuelle stratigraphique entre les métadonnées de Dees et les métadonnées philologiques. L'ensemble des configurations présence/absence des attributs est représenté en table 2.3 pour les 296 textes du NCA.

2.1.4.1 Métadonnées : la place de l'Atlas de 1987 dans le NCA

Les 200 textes de l'Atlas de 1987 correspondent en majorité à des textes dotés de la métadonnée regionDees. Mais cette correspondance n'est pas parfaite : seuls 198 textes disposent d'une regionDees. Par ailleurs, comme évoqué *supra*, un texte de l'Atlas est manquant dans le NCA (ren1), et un des textes de Dees est dédoublé dans le NCA (Aucassin et Nicolette). Il manque donc deux textes. L'un peut être aisément identifié : hista. Ce texte dispose d'un coefficient dialectal de 75 et d'un coderegional de 61, lequel correspond à la Haute-Marne, et les informations bibliographiques données par le NCA pour hista concordent avec celles données par Dees dans sa bibliographie pour le texte haut-marnais *Li livres des estoires Rogier*. Si nous ne connaissons pas la raison pour laquelle certaines balises du XML de ce texte ont pour attribut NIL, y compris celles relatives à la bibliographie de l'Atlas (dont la regionDees), son identification semble assez probable.

L'autre texte manquant est une version normande du *Roman de Rou* de Wace et, ici, l'identification du texte correspondant au sein du NCA impose de supposer une erreur manifeste lors de la saisie des données. La correspondance entre ce *Roman de Rou* normand de l'Atlas et le texte rou3b semble pourtant évidente : l'Atlas répertorie deux *Romans de Rou*, l'un anglo-normand (coef. dial. = 70), et l'autre normand (coef. dial. =

TABLE 2.3 – Combinaisons des métadonnées du NCA3.

'✓' représente un attribut non nul, '×' représente l'attribut nil, équivalent de NA dans les fichiers XML du NCA3.

region Dees	code Regional	date Moyenne Dees	date Composition	date Manuscrit	lieu Composition	lieu Manuscrit	Nb.
✓	✓	✓	✓	✓	✓	✓	129
✓	✓	✓	✓	✓	✓	×	7
✓	✓	✓	✓	✓	×	✓	19
✓	✓	✓	✓	✓	×	×	4
✓	✓	✓	✓	×	×	×	1
✓	✓	×	✓	✓	✓	✓	28
✓	✓	×	✓	✓	✓	×	4
✓	✓	×	✓	✓	×	✓	2
✓	✓	×	✓	✓	×	×	1
✓	✓	×	✓	×	✓	×	1
✓	✓	×	×	✓	×	✓	1
✓	×	✓	✓	✓	×	×	1
×	✓	✓	✓	✓	✓	✓	21
×	✓	✓	✓	✓	✓	×	2
×	✓	✓	✓	✓	×	✓	4
×	✓	×	✓	✓	✓	✓	31
×	✓	×	×	×	×	×	1
×	×	✓	×	✓	×	×	1
×	×	×	✓	✓	✓	✓	31
×	×	×	✓	✓	✓	×	4
×	×	×	✓	✓	×	✓	1
×	×	×	✓	×	×	×	2

82). Le texte rou3a du NCA est effectivement donné comme se rattachant à l'Angleterre avec un coefficient de 70, tandis que la reconstruction de la référence bibliographique de rou3b par Glessgen et son équipe est cohérente avec celle du *Rou* normand de la bibliographie de l'*Atlas* de 1987. En revanche, le coefficient de dialectalité du texte, tel qu'indiqué dans le préambule du XML, est de 56, alors que Dees donne, dans sa bibliographie, un coefficient de 82. Enfin, le codeRegional de ce texte est de 86, ce qui renvoie à l'Angleterre. Les indications de localisation philologiques penchent, elles aussi, vers la Normandie (tant pour la composition que pour la copie). Aussi, si la bibliographie restituée par l'équipe du NCA et la présence de deux *Romans de Rou* dans l'*Atlas* laissent supposer que le roman manquant correspond à rou3b, le coefficient de dialectalité et le code régional des métadonnées s'y opposent. Le décalage entre les textes dotés d'une regionDees et ceux de l'*Atlas* se réduit ainsi à la disparition de ren1, à la lacune de la métadonnée regionDees dans le préambule de hist a et au sort du *Roman de Rou* Normand, correspondant probablement à rou3a

Par ailleurs, 35 textes sont dotés d'un coefficientRegional inférieur à 70, allant de 52 à 69. Ils correspondent selon toute probabilité à la catégorie de textes (b), c'est-à-dire ceux qui étaient destinés à l'*Atlas* mais on été exclus sur la base de leur seul coefficient de dialectalité, à l'exception de rou3b si celui-ci dispose de métadonnées erronées. Enfin, 62 textes n'ont

pas de `coefficientRegional` ; ils n'ont pas non plus de `regionDees`. Sur ces 62 textes, 23 disposent pourtant d'un `codeRegional` correspondant à l'Angleterre et, pour ces 23 textes, la localisation philologique proposée par Glessgen et son équipe est cohérente avec cela, mais avec un degré de précision variable : 11 textes sont donnés comme ayant été composés en anglo-normand,²¹ 11 autres textes sont donnés comme ayant simplement été composés dans le « Nord-Ouest »,²² ce qui inclut donc l'Angleterre et l'aire dialectale anglo-normande, et un texte n'a pas de composition localisée (`oakbook`). Pour ce dernier, le titre indique clairement une origine anglo-normande : le texte a été édité comme *Ancient ordinances of the gild merchant of the town of Southampton* (SMIRKE 1859). Les localisations de copie, en revanche, sont à la fois plus précises et plus divergentes : 19 textes sont de copie anglo-normande ou anglaise,²³ 1 texte est de copie normanno-anglo-normande (`fablesY`) mais un texte est de copie wallone (`fablesM`) et un est de copie champenoise (`fablesT`). On peut néanmoins comprendre le principe de rattachement des ces trois derniers textes, dont lieu de composition et lieu de copie différent, par leur autrice commune : ce sont trois versions des *Fables* de Marie de France.

Enfin, les 39 textes restant disposent de localisations variées. Ils ont principalement été composés dans un dialecte du Nord-Ouest du domaine,²⁴ en champenois,²⁵ à Tournai (`fierens`), en dialecte dit francien (`lechS`), picard (`songe14`) ou indéfini.²⁶ Les manuscrits, eux, sont principalement franciens (13 textes),²⁷ picards (7 textes),²⁸ et anglo-normands (4 textes);²⁹ deux sont champenois (`pera`, `yvh`), un est bourguignon (`perb`), un est lorrain (`fablesC`), un est normand (`narcD`), un provient de Tournai (`fierens`), un provient d'un dialecte indéfini du Nord-Est (`perpraag`), et sept n'ont pas de lieu ou de dialecte de copie précisé.³⁰

2.1.4.2 Distribution diachronique des textes

Les textes du corpus peuvent être envisagés selon trois métadonnées diachroniques : la date indiquée par Dees (`dateMoyenneDees`), la date de composition indiquée par le travail philologique (`dateComposition`, dorénavant *compo.*) et la date de copie issue également du travail philologique (`dateManuscrit`, désormais *ms.*). Selon chacune de ces trois informations, le corpus du NCA s'étend de 1150 à 1450 (`dateMoyenneDees`), de 1100 à 1489 (*compo.*), ou de 1137 à 1489 (*ms.*), avec un texte isolé dont le *ms.* date de 1750 (*verite*). Le corpus utilisé dans l'*Atlas* de 1987 s'étend, selon ces informations, de 1150 à 1400 (`dateMoyenneDees`), de 1125 à 1365 (*compo.*) ou de 1190 à 1400 (*ms.*). Ces trois manières de concevoir la composition diachronique du corpus sont représentées en fig. 2.5.

L'usage des dates de composition (fig. 2.5 b) tend à anticiper le centre de gravité du corpus, le déplaçant vers le début du XIII^e s., tandis que l'usage des dates de copie (c) déplace ce centre de gravité en direction de la fin du siècle. Par ailleurs, pour la production de ces graphiques comme pour le reste de l'analyse de ce travail, nous avons dû normaliser les dates : dans le cas où une date est indiquée avec une fouchette ou un *circa*, nous avons simplement ignoré cet indice.

Le dessin en dents de scie observable dans les `dateMoyennesDees` (a) est probablement dû à un effet analogue : les textes étant régulièrement datés en fonction des siècles, de leur début ou de leur fin, une normalisation

21: `chaitH`, `chevreH`, `deusamH`, `edmond`, `edmund`, `equiH`, `ipo`, `plainte`, `psautier`, `reis` et `yonecH`.

22: `fablesA`, `fablesB`, `fablesE1`, `fablesM`, `fablesT`, `fablesY`, `guigH`, `lanvalC`, `laustH`, `milunH`, `rolandox`.

23: `chaitH`, `chevreH`, `deusamH`, `edmond`, `edmund`, `equiH`, `fablesA`, `fablesB`, `fablesE1`, `guigH`, `ipo`, `lanvalC`, `laustH`, `milunH`, `plainte`, `psautier`, `reis`, `rolandox` et `yonecH`.

24: Nord-Ouest (10) : `fablesC`, `fablesK`, `fablesL`, `fablesZ`, `guigP`, `guigS`, `lanvalH`, `lanvalP`, `lanvalS` et `milonS`.

Anglo-normand (9) : `chevreS`, `deusamS`, `equiS`, `prologueH`, `sergbH`, `sergb0`, `yonecP`, `yonecQ` et `yonecS`.

Normand (7) : `narcA`, `narcB`, `narcC`, `narcD`, `narcE`, `rou1` et `rou2`.

25: `pera`, `perb`, `perpraag`, `perr`, `yvf`, `yvh`.

26: `rombriv`, `rombriva`, `romi` et `verite`.

27: `chevreS`, `deusamS`, `equiS`, `guigS`, `lanvalS`, `lechS`, `milonS`, `narcA`, `narcB`, `rou1`, `rou2`, `yonecQ` et `yonecS`.

28: `guigP`, `lanvalP`, `narcC`, `rombriv`, `yonecP`, `yvf` et `perr`.

29: `lanvalH`, `prologueH`, `sergbH`, `sergb0`.

30: `fablesK`, `fablesL`, `fablesZ`, `rombriva`, `romi`, `songe14` et `verite`.

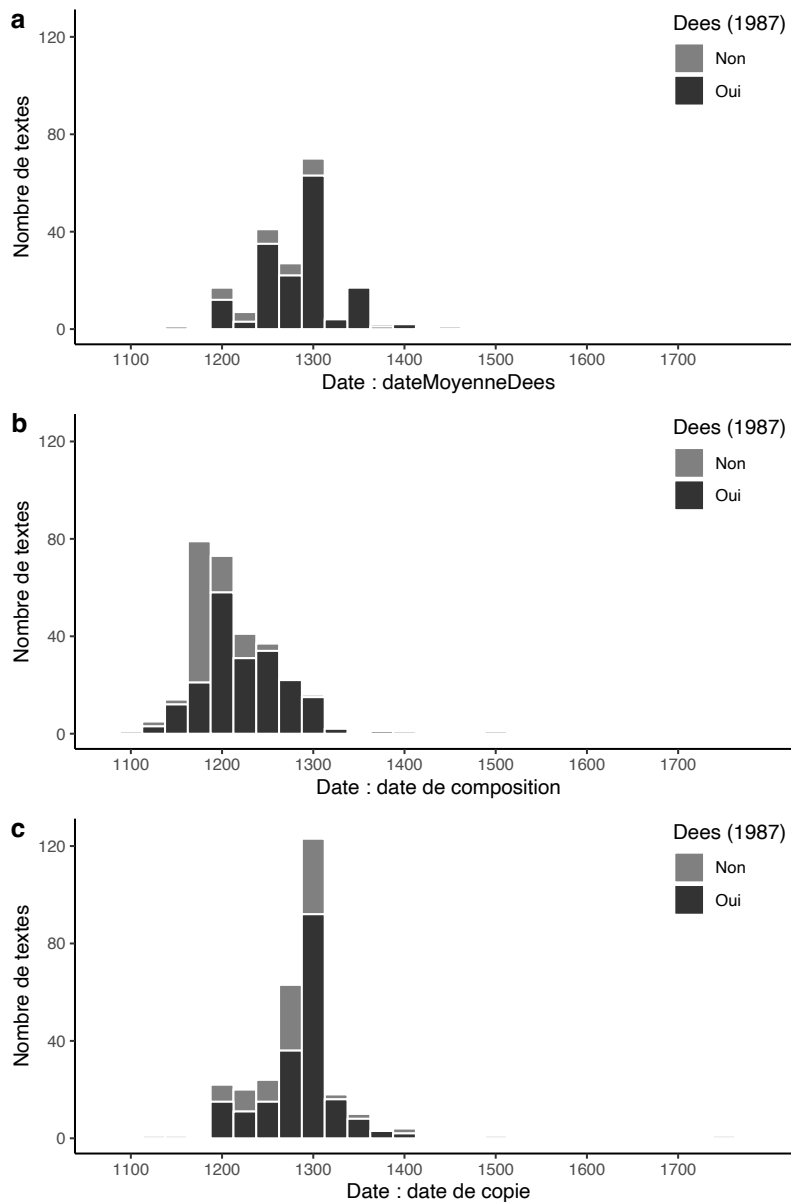


FIGURE 2.5 – Distribution des textes du NCA en fonction des trois datations disponibles.

sous format numérique produit des pics au changement des siècles (fin de siècle et début de siècle inclus) et au milieu de ceux-ci (lorsqu'un texte est daté soit d'un siècle sans plus de précisions soit du milieu d'un siècle). Le travail philologique de Glessgen et son équipe donnant des dates auxquelles des précisions sont rajoutées, cet effet en dents de scie n'est pas représenté en (b) et (c) : même si la précision est ignorée, ce travail a moins tendance à placer systématiquement les textes en début, milieu et fin de siècle, s'appuyant plutôt sur une précision apportée à une date numérique. Enfin, en (b) et (c), le nombre de textes est plus important (b : 293 textes datés ; c : 291 textes datés), étant donné que les `dateMoyenneDees` manquent non seulement sur une partie du corpus de l'*Atlas* de 1987 (40 textes) mais également sur la majorité du corpus exclu

de l'*Atlas* (67 textes). De ce fait, l'utilisation des dates Moyennes Dees n'est pas viable pour un travail systématique, y compris en se restreignant au seul corpus de l'*Atlas*.

Quelle que soit la donnée retenue, le NCA couvre convenablement l'ensemble du XIII^e s. Les différences entre les trois datations tiennent au fait que l'usage des dates de composition permet une couverture convenable d'au moins la seconde moitié du XIII^e s., ce que n'autorisent ni les dates Moyennes Dees ni les dates de copie, tandis que les dates de copie permettent, en miroir, une couverture convenable de la première moitié du XIV^e s. Il convient par ailleurs de noter que les dates de copie impliquent un corpus moins bien équilibré que les deux autres datations, avec un pic très marqué autour de 1300, tandis que les deux autres datations proposent des corpus plus lissés. Enfin, les dates de composition et de copie comportent des *outliers*, c'est-à-dire des textes fortement détachés du reste du corpus. Du point de vue des dates de composition, ces outliers sont loth (1365), songe14 (1389) et fierens (1489). Du point de vue des dates de copie, ce sont fierens (1489) et verite (1750). L'écart entre le premier de ces textes et le texte qui les précède, 43 ans pour les dates de composition et 89 ans pour les dates de copie, est inédit dans les corpus : tous les autres textes sont séparés par un écart plus faible (25 ans entre la composition de rolandox et celle de psautier, 40 ans entre la copie de ces deux mêmes textes).³¹ On peut donc, sans avoir besoin de plus d'investigations à propos de ces textes, les écarter du corpus étant donné qu'il se placent, d'eux-mêmes, à l'écart de celui-ci et sont trop peu nombreux pour assurer une quelconque représentativité de leur époque.

2.1.4.3 Distribution diatopique des textes

Les textes du corpus peuvent également être envisagés selon trois métadonnées diatopiques : la localisation dialectométrique de Dees (`regionDees`), la localisation de composition indiquée par le travail philologique (`lieuComposition`) et la localisation de copie issue également de la philologie (`lieuManuscrit`). Comme pour les dates, les localisations philologiques du NCA sont hétérogènes : d'une part elles mélangent points de monde physique (p.ex. « La Rochelle », « Oise », « Belgique ») et dialectes (p.ex. « frc » [= francien], « agn » [= anglo-normand]), et d'autre part leur granularité est irrégulière, du grain le plus fin (p.ex. « Liège », « La Rochelle ») au grain le plus grossier (p.ex. « Ouest », « Nord-Ouest »). Dans certains cas, l'interaction entre ces deux sources d'hétérogénéité donne un attribut dialectal qui ne saurait correspondre à une localisation (p.ex. `myst : lieuComposition` = « traits occ. et agn. » [= traits occitans et anglo-normands]). Ces localisations philologiques, si elles sont riches en renseignements pour un lecteur humain, ne sont pas destinées à un traitement informatique, et sont difficiles à standardiser à cette fin. Sur le plan pratique, les localisations de Dees sont donc clairement favorisées par leur facilité d'usage.

Sur le plan méthodologique, les localisations de Dees disposent aussi de l'avantage d'une méthode systématique, ne faisant jamais appel à l'intuition ou à l'érudition du philologue. Si, dans la plupart des cas, les localisations de Dees concordent plus ou moins avec au moins l'une des deux localisations philologiques, il est des textes pour lesquels le

31: Distance entre deux textes successifs :

Date de composition :

- Moy. = 0,77 ans,
- Max. = 25 ans,
- σ = 2,23.

Date de copie :

- Moy. = 0,9 ans,
- Max. = 40 ans,
- σ = 3,48.

Dans les deux cas, la médiane est de 0, étant donné que la majorité des textes dispose d'une date identique à celle d'au moins un autre texte (218 textes pour la date de composition, 244 pour la date de copie).

désaccord est patent (cf. GLEÛGEN et GOUVERT 2007, p. 57-58). Plutôt que de trancher en faveur de la philologie lorsque celle-ci présente un « haut degré de certitude » comme le proposent GLEÛGEN et GOUVERT (2007, p. 47), nous préférons considérer ici qu'en cas de désaccord chaque approche documente un aspect différent du texte. En effet, l'approche de Dees ne s'intéresse qu'à la langue du texte, réduite à 268 traits, tandis que l'approche philologique considère également l'histoire du texte et de son ou ses supports, et dissèque la langue du texte en strates isolables. La dissection la plus commune, et directement représentée dans les entrées du NCA, est celle entre composition et copie finale, qui s'applique à tous les textes dont la composition n'est pas contemporaine de la copie finale. Mais l'approche philologique amène aussi à isoler des caractéristiques hétérogènes au sein de la même strate chronologique : ainsi des *traits* dialectaux au sein de textes qui ne correspondraient pas exactement à cette origine dialectale (2.3 a-g) et des indications de plusieurs origines dialectales co-existantes (2.3 h-k).

(2.3) Localisations philologiques par traits

(a)	robert	compo. ms. Dees	traits norm. ? pic. Somme et Pas-de-Calais (75 : Pas-de-Calais centre + nord)
(b)	meun	compo. ms. Dees	traits orl. orl. Région parisienne (89 : Seine-et-Marne)
(c)	rose	compo. ms. Dees	traits orl. orl. Nièvre et Allier (84 : Nièvre et Allier)
(d)	sage	compo. ms. Dees	traits de l'Ouest frc. Région parisienne (90 : Seine-et-Marne)
(e)	fables ^Y	compo. ms. Dees	Nord-Ouest norm. et agn. Angleterre (86) ³²
(f)	pers	compo. ms. Dees	champ. mérid. traits norm. Normandie (89 : Eure)
(g)	kathe	compo. ms. Dees	poit. traits frpr. Franche-Comté (75 : Haute-Saône)
(h)	alia	compo. ms. Dees	pic. et occ. orient. poit. Vendée et Deux-Sèvres (71 : Vendée)
(i)	myst	compo. ms. Dees	traits occ. et agn. tour. Angleterre (57)
(j)	aye	compo. ms. Dees	pic./norm. nil Normandie (62) ³³
(k)	loys	compo. ms. Dees	multiple nil Normandie (96 : Eure)

32: Les textes anglo-normands n'ont jamais de sous-région.

33: Pas de sous-région indiquée.

Dans certains cas, lorsque la localisation philologique ne mentionne que des traits ou des caractéristiques plurielles, Dees lui-même ne parvient pas à un coefficient de dialectalité satisfaisant (2.3 i, j). Néanmoins, il obtient un coefficient supérieur à 70 dans la plupart des cas, et même égal ou supérieur à 89 dans quatre cas (2.3 b, d, f, k). Ainsi, alors que la localisation philologique s'en tient à mentionner des traits, la méthode dialectométrique parvient, par la quantification de ceux-ci, à définir si la présence de ces traits est suffisante pour déclencher l'association d'une localisation, ou si ces traits sont trop marginaux. De cette manière, pour certains textes, les traits indiqués ne parviennent pas à contrer la disparition du caractère dialectal (2.3 b, d); pour d'autres, ce que la bibliographie note comme de simples traits constitue un marquage dialectal suffisant fort (2.3 f, g,³⁴ i, k), ou la présence de traits discordants ne suffit pas à annuler la forte prédominance d'un autre dialecte (2.3 c, e). Enfin, et parallèlement, lorsque la localisation de Dees et/ou les localisations philologiques concluent à une langue se rapportant à la région parisienne, il faut entendre cette localisation comme étant plus abstraite que les autres : ce que la philologie qualifie de *francien* est avant tout un affaiblissement du caractère dialectal du texte, une « pré-standardisation » qui se joue plus sur la sociolinguistique de la langue employée que sur sa seule diatopie (GLEÏGEN et GOUVERT 2007, p. 60-61), loin de l'unité dialectale de l'Île-de-France à laquelle certains ont voulu croire (BERGOUNIOUX 1989).

34: Si l'on considère qu'il y a une correspondance, même imparfaite, entre ce que la bibliographie identifie comme des traits francoprovençaux et le dialecte franc-comtois.

BERGOUNIOUX (1989). « Le francien (1815-1914) : la linguistique au service de la patrie ».

2.1.5 Choix des métadonnées utilisées

Dans ce travail, chaque fois que nous interrogeons le NCA, nous documentons la distribution diatopique et diachronique des données. Comme exposé *supra*, plusieurs métadonnées permettent de documenter ces dimensions, et le choix des métadonnées à utiliser est non seulement un choix pragmatique, mais également scientifique. Dans ce qui suit, nous expliquons les choix que nous avons effectués parmi ces métadonnées.

2.1.5.1 Métadonnées diachroniques

Parmi les trois dates qui sont disponibles, à savoir la datation de Dees, la date de composition et la date de copie, la première se distingue des deux autres, sur plusieurs aspects. D'une part, elle est normalisée, et se prête à un traitement informatique simple : elle se résume toujours à un nombre, tandis que les dates d'origine philologique sont augmentées de *circa* ou de fourchettes temporelles. D'autre part, elle n'est pas documentée : nous ne savons pas sur quelles sources Dees s'est appuyé pour établir cette date ; alors que la source des datations philologiques est toujours indiquée dans le préambule du texte. Enfin, la datation de Dees manque pour 107 textes, soit un peu plus d'un tiers du corpus. Ainsi, même si son usage informatique est plus aisé, l'absence de documentation des sources et la défection sur plus d'un tiers des textes nous a amené à ignorer la datation de Dees, au profit des datations philologiques.

Entre les datations philologiques, nous avons préféré retenir la datation de composition, au détriment de la datation de copie. S'il serait envisageable d'examiner chaque résultat du corpus en fonction des deux dates

philologiques disponibles, il s'agit là d'un travail considérable, qui n'est pas l'objet de cette thèse. Si nous avons préféré la date de composition à la date de copie, c'est en bonne partie grâce au travail exploratoire réalisé sur le traitement des schwas internes, qui montre une meilleure cohérence diachronique en fonction de la date de composition qu'en fonction de la date de copie (cf. chap. 11). Sur ce fondement, il nous a semblé que c'était la date de composition qui représentait le mieux la chronologie dans laquelle se jouent les chutes de voyelles; ce qui ne dit rien de sa capacité à représenter efficacement d'autres variations phonologiques.

Par ailleurs, les dates de composition et de copie ont dû être normalisées, par simple effacement d'un éventuel *circa* ou d'une fourchette temporelle. Le degré de précision supposée de l'information a donc été perdu, ce qui invite à considérer avec prudence les informations diachroniques : elles sont aptes à représenter une dynamique d'ensemble, mais on ne saurait leur prêter une grande fidélité à l'échelle, par exemple, d'une décennie. Enfin, pour les trois textes ne disposant pas de date de composition, à savoir *arr*, *oakbook* et *verite*, nous avons choisi de ne pas suppléer à l'absence de date de composition par la date de copie, d'une part pour ne pas introduire d'hétérogénéité dans cette variable, d'autre part parce que, sur ces trois textes, l'un ne dispose pas non plus d'une date de copie (*oakbook*), tandis qu'un autre dispose d'une date de copie extrêmement tardive (*verite* : 1750). Aussi, introduire une supplétion par les dates de copie nous a paru superflu, sachant que cette supplétion ne concernerait, en définitive, qu'un seul texte. La distribution diachronique du corpus, tel que nous l'explorons, est donc celle reproduite en fig. 2.5 b, *supra*.

2.1.5.2 Métadonnées diatopiques

Comme pour la datation, l'indication géographique de Dees se distingue de celles de issues du travail philologique, et ce à plusieurs titres. Sur le plan pratique, si elle contient parfois un empilement de données par l'intermédiaire des sous-régions, elle contient toujours un niveau constant et discret, celui des régions de Dees. Ces régions, si elles correspondent à un espace dialectologique projeté sur l'espace géographique, sont définies explicitement par les cartes que Dees a produites, et sont reproductibles grâce aux *shapes files* réalisés par Guylaine Brun-Trigaud pour l'ADE22. Ce sont ces fichiers spatiaux qui nous permettent de projeter sur une carte les données du corpus.

Au contraire, les indications géographiques issues de la philologie ne comportent pas de niveau commun à l'ensemble des textes localisés et sont hétérogènes par nature, allant de l'indication dialectale à l'indication géographique, des traits mélangés à une ville précise. Par ailleurs, sur le plan de la méthode, les localisations de Dees présentent l'avantage de localiser la langue du texte comme un ensemble unitaire, ce qui est essentiel pour placer celui-ci sur une carte. Enfin, la méthode de Dees assure une localisation homogène entre les textes, insensible aux angles morts ou aux contradictions de la littérature. Nous avons donc privilégié la localisation fournie par Dees.

Néanmoins, la localisation de Dees manque, sous sa forme *regionDees*, pour 98 textes. Nous avons mis en place un processus de supplétion

systématique pour attribuer une région à ces 98 textes. D’abord, pour 58 textes, l’attribut de `regionDees` est nil, mais l’information est présente sous la forme du `codeRegional`, et peut donc être aisément restituée. Dans ce cas, il ne s’agit que de corriger une lacune en rétablissant une redondance dans les métadonnées, et la localisation est bien issue du travail de Dees. Il reste donc 40 textes pour lesquels nous n’avons aucune localisation issue du travail de Dees. La philologie propose une localisation de composition pour 36 de ces textes, et une localisation de copie pour 34 de ces textes. Nous avons restitué une `regionDees` sur le fondement de ces localisations philologiques, en privilégiant la localisation de composition lorsqu’elle est disponible et qu’elle est suffisamment précise; la localisation de copie n’est utilisée que lorsque la localisation de composition est trop imprécise (`lanvalH`, `fablesC`, `guigS`, `lanvalS`, `milonS`, `guigP`, `lanvalP`) ou manquante (`rombriv`). Ce travail a nécessité une normalisation des localisations philologiques, dont le détail est donné en table 2.4, page suivante.

Cette normalisation a notamment impliqué d’effacer l’information « méridional » et « septentrional » de l’étiquette champenoise, dans la mesure où certains textes sont uniquement qualifiés de champenois et ne peuvent donc pas s’insérer dans cette répartition Nord/Sud de la Champagne. Si l’attribution de correspondances entre les `regionDees` et les localisations philologiques est parfois transparente (Angleterre \Leftrightarrow agn.; Normandie \Leftrightarrow norm.; Région parisienne \Leftrightarrow frc.) ou simple à déduire (dans le cas de villes : Troyes \Rightarrow Aube; Tournai \Rightarrow Wallonie), elle est moins transparente pour d’autres régions, pour lesquelles des choix ont dû être faits. Ainsi, nous avons choisi de placer dans la `regionDees` de l’Aube tous les textes champenois, alors que la Champagne historique et que le dialecte champenois ne sauraient se réduire à ce seul département moderne. Il en va de même pour l’étiquette picarde, qui a été mise en correspondance avec la `regionDees` Somme, Pas-de-Calais, ainsi que pour l’étiquette lorraine, qui a été mise en correspondance avec la `regionDees` Moselle, Meurthe-et-Moselle.

Cette méthode pose évidemment un certain nombre de problèmes. Certains problèmes tiennent simplement à la normalisation appliquée. D’une part, elle transforme en localisations discrètes des localisations qui ne le sont pas nécessairement, et écrase en un seul type d’information des informations géographiques et dialectales telles que fournies par la bibliographie philologique. D’autre part, elle force des associations entre des entités historiques ou dialectales et des départements modernes, pour la champagne par exemple. Enfin, elle recourt à deux sources différentes : elle ignore la localisation du ms. lorsque la localisation de composition est disponible et satisfaisante, mais recourt à la localisation du ms. lorsque la localisation de composition est absente ou impossible à mettre en correspondance avec une `regionDees`.

D’autres problèmes, et ils ne sont pas moins sérieux, tiennent aux effets de cette méthode : non seulement cette méthode ne fait aucun usage des coefficients de dialectalité de Dees (ainsi des 35 textes aux scores inférieurs à 70), mais elle crée de nouvelles localisations que l’on ne peut mesurer à l’aune de cet indicateur. Il faut néanmoins remettre cette question en perspective : dans le projet d’atlas qui est celui de Dees, la localisation est évidemment l’élément le plus important, et il peut se permettre d’écarter des textes à la localisation insatisfaisante. Mais un

TABLE 2.4 – NCA : obtention d'une *regionDees* à partir des localisations philologiques. Les valeurs entre parenthèses sont celles qui n'ont pas été retenues pour l'obtention d'une correspondance avec les *regionDees*.

id	regionDees_supp	lieuComposition	lieuManuscrit
chevreS	Angleterre	⇐ agn.	(frc.)
deusamS	Angleterre	⇐ agn.	(frc.)
equiS	Angleterre	⇐ agn.	(frc.)
lanvalH	Angleterre	⇐ (Nord-Ouest)	⇐ agn.
prologueH	Angleterre	⇐ agn.	(agn.)
sergbH	Angleterre	⇐ agn.	(agn.)
sergb0	Angleterre	⇐ agn.	(agn.)
yonecP	Angleterre	⇐ agn.	(pic.)
yonecQ	Angleterre	⇐ agn.	(frc.)
yonecS	Angleterre	⇐ agn.	(frc.)
pera	Aube	⇐ champ. merid.	(champ.)
perb	Aube	⇐ champ. merid.	(bourg. sept.)
perpraag	Aube	⇐ Troyes	(Nord-Est)
perr	Aube	⇐ champ. merid.	(pic.)
yvf	Aube	⇐ champ. merid.	(pic.)
yvh	Aube	⇐ champ. merid.	(champ.)
fablesC	Moselle, Meurthe-et-Moselle	⇐ (Nord-Ouest)	⇐ lorr.
narcA	Normandie	⇐ norm.	(frc.)
narcB	Normandie	⇐ norm.	(frc.)
narcC	Normandie	⇐ norm.	(pic.)
narcD	Normandie	⇐ norm.	(norm.)
narcE	Normandie	⇐ norm.	(art.)
rou1	Normandie	⇐ norm.	(frc.)
rou2	Normandie	⇐ norm.	(frc.)
guigS	Region parisienne	⇐ (Nord-Ouest)	⇐ frc.
lanvalS	Region parisienne	⇐ (Nord-Ouest)	⇐ frc.
lechS	Region parisienne	⇐ frc.	(frc.)
milonS	Region parisienne	⇐ (Nord-Ouest)	⇐ frc.
guigP	Somme, Pas-de-Calais	⇐ (Nord-Ouest)	⇐ pic.
lanvalP	Somme, Pas-de-Calais	⇐ (Nord-Ouest)	⇐ pic.
rombriv	Somme, Pas-de-Calais	⇐ (nil)	⇐ pic.
fierens	Wallonie	⇐ Tournai	(Tournai)
fablesL	NA	(Nord-Ouest)	nil
fablesK	NA	(Nord-Ouest)	nil
fablesZ	NA	(Nord-Ouest)	nil
livre	NA	(Ouest)	(Ouest)
rombriva	NA	nil	nil
romi	NA	nil	nil
verite	NA	nil	nil

atlas est un outil général, qui ne permet pas de documenter de manière satisfaisante toutes les facettes de tous les phénomènes. Notre projet, ici, est donc différent : nous ne visons à documenter qu'un seul phénomène, le devenir des voyelles finales atones, et ne visons pas à écraser un siècle et demi de diachronie en une synchronie globale. Or, ce projet, contrairement à celui d'un atlas général, active d'autres enjeux : non seulement tous les textes ne répondront pas à nos interrogations, mais nous avons également besoin d'avoir autant de textes que possible, dans toutes les régions, à tout instant de la diachronie (ou, au moins, disons, dans chaque demi-siècle). Enfin, tout en étant problématique en théorie, cette méthode ne semble pas l'être, en pratique, pour certains textes : ainsi de *lanvalH*, *prologueH*, *sergbH*, *sergb0*, *pera*, *yvh*, *narcD*, *lechs* et *fierens*, pour lesquels la localisation de composition est cohérente avec la localisation de copie.

L'enrichissement des données géographiques par cette méthode est illustré par la comparaison des cartes en fig. 2.6 (*regionDees* uniquement), 2.7 (*codeRegional*) et 2.8 (*regionDees*, complétées par *codeRegional* et par les localisations philologiques). La comparaison des deux premières cartes avec la troisième carte montre que l'utilisation des localisations philologiques ne bouleverse pas fondamentalement le paysage : si elle renforce sensiblement le poids du Nord-Ouest, avec l'Angleterre (8 textes en *regionDees* contre 48 après complétion par les données philologiques), la Normandie (15 textes contre 25) et la région Somme, Pas-de-Calais (37 textes contre 45), elle n'améliore que légèrement le poids du Sud-Ouest pris dans son ensemble, en proposant 7 textes de plus dans la région Vendée, Deux-Sèvres (5 textes contre 12), et ne renforce que ponctuellement celui de l'Est, principalement dans l'Aube (9 textes contre 16) et la Wallonie (11 textes contre 14), ainsi qu'en la région parisienne (14 textes contre 18). En écrasant le corpus en synchronie, l'intérêt d'une telle méthode n'est donc guère palpable. C'est dans la prise en compte de la diachronie que cette méthode révèle ses bénéfices. En fig. 2.9 et 2.10, nous avons divisé le corpus en quatre périodes, selon des demi-siècles, et le nombre de textes est donné, pour chaque région, selon chaque type de localisation présente dans les métadonnées du corpus.

1100-1150. Comme illustré précédemment pour la distribution diachronique du corpus, celui-ci ne dispose que de peu de textes antérieurs à 1151. Ici, ce n'est pas l'application de nos correspondances entre localisations philologiques et régions de Dees qui change quoi que ce soit ; en revanche, l'utilisation des codes régionaux de Dees, au lieu de l'utilisation des noms de régions, améliore sensiblement la situation de l'Angleterre (qui passe de 1 à 4 textes) et, à la marge, celle de la Haute-Marne (de 3 à 4 textes) (fig. 2.9 a). Ces différences sont simplement dues au fait que les trois textes anglais (*rolandox*, *psautier*, *teo2*) et le texte de la Haute-Marne (*mrgr1*) ont été pré-traités par Dees mais écartés par la suite (aucun n'a de coefficient de dialectalité ou de sous-région).

1151-1200. Dans la seconde moitié du XII^e s. (fig. 2.9 b), l'effet du choix de la variable géographique devient plus fort : sur 132 textes, l'information *regionDees* n'est présente que pour 58 textes, alors que l'information *codeRegional* est présente pour 95 textes, et qu'on peut atteindre 126 textes en recourant aux localisations philologiques. La différence la plus

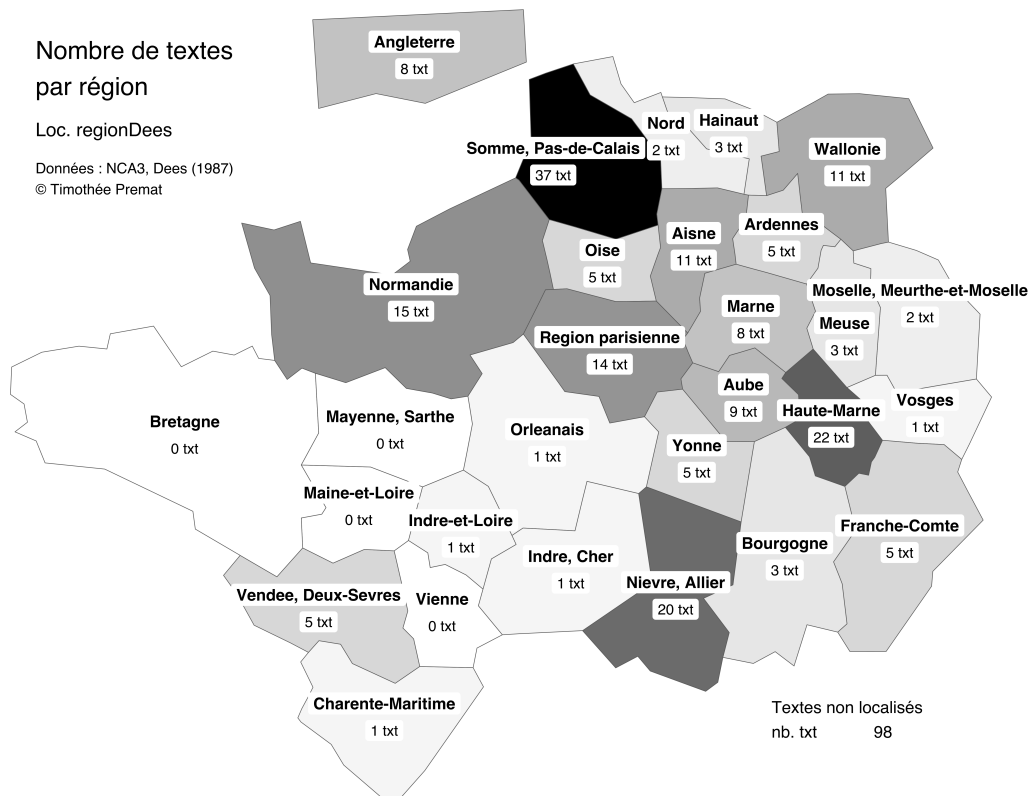


FIGURE 2.6 – Répartition des textes du NCA selon les localisations regionDees.

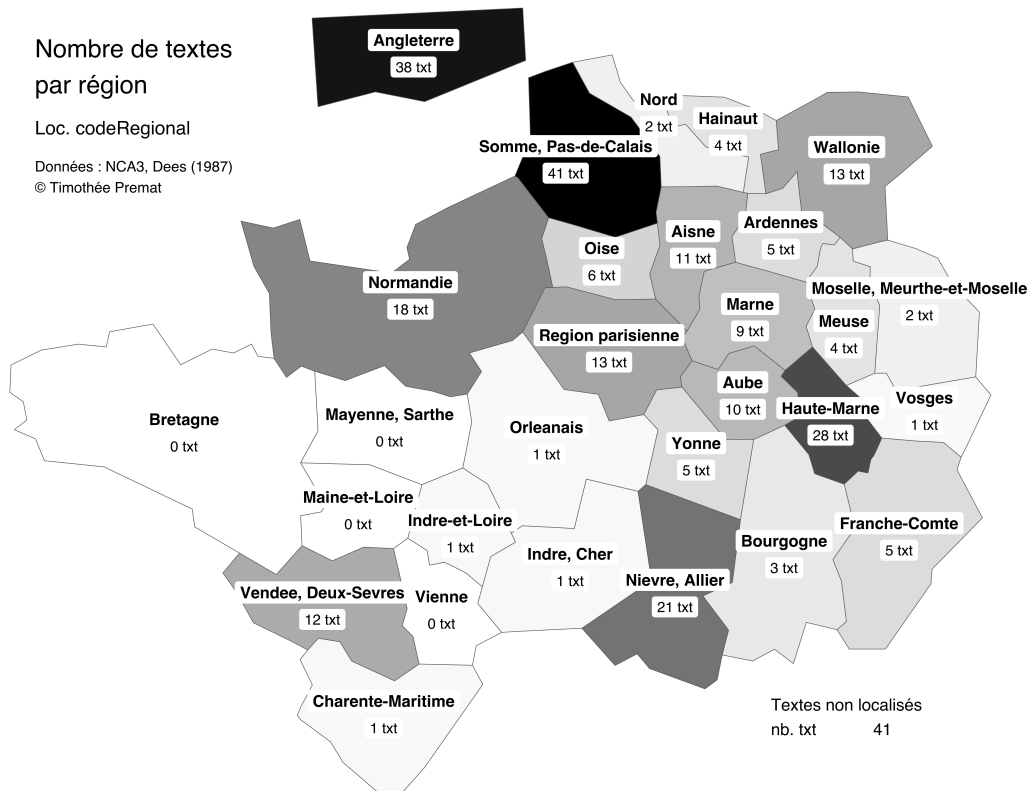


FIGURE 2.7 – Répartition des textes du NCA selon les localisations codeRegional.

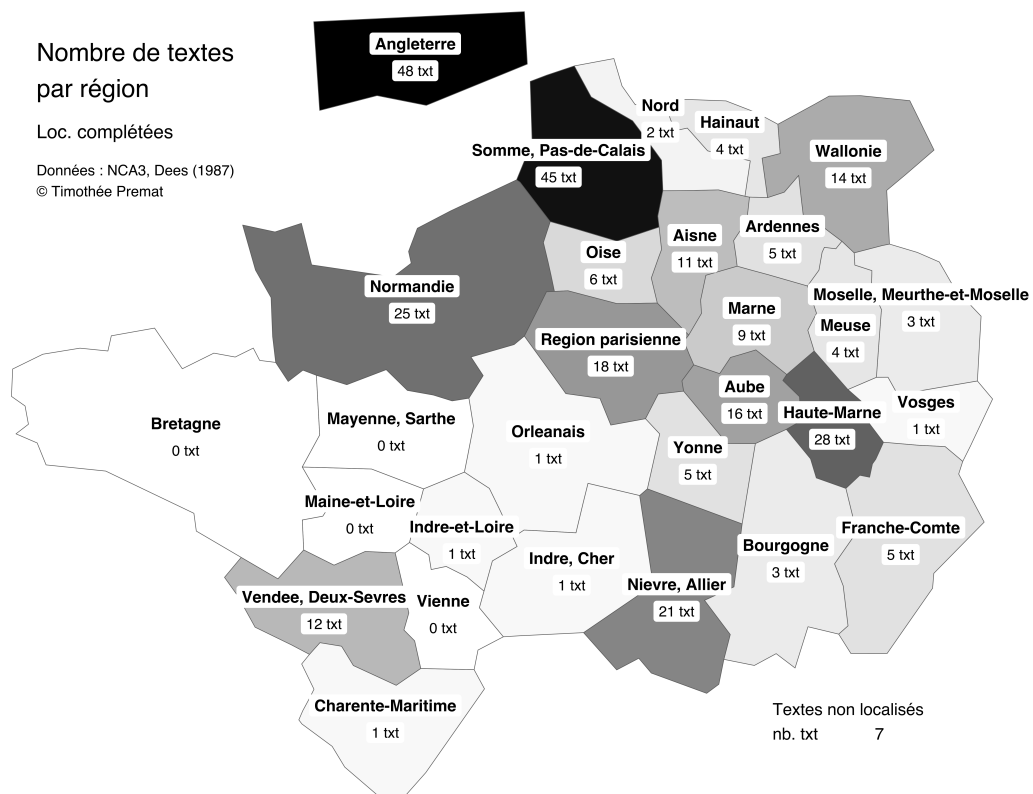
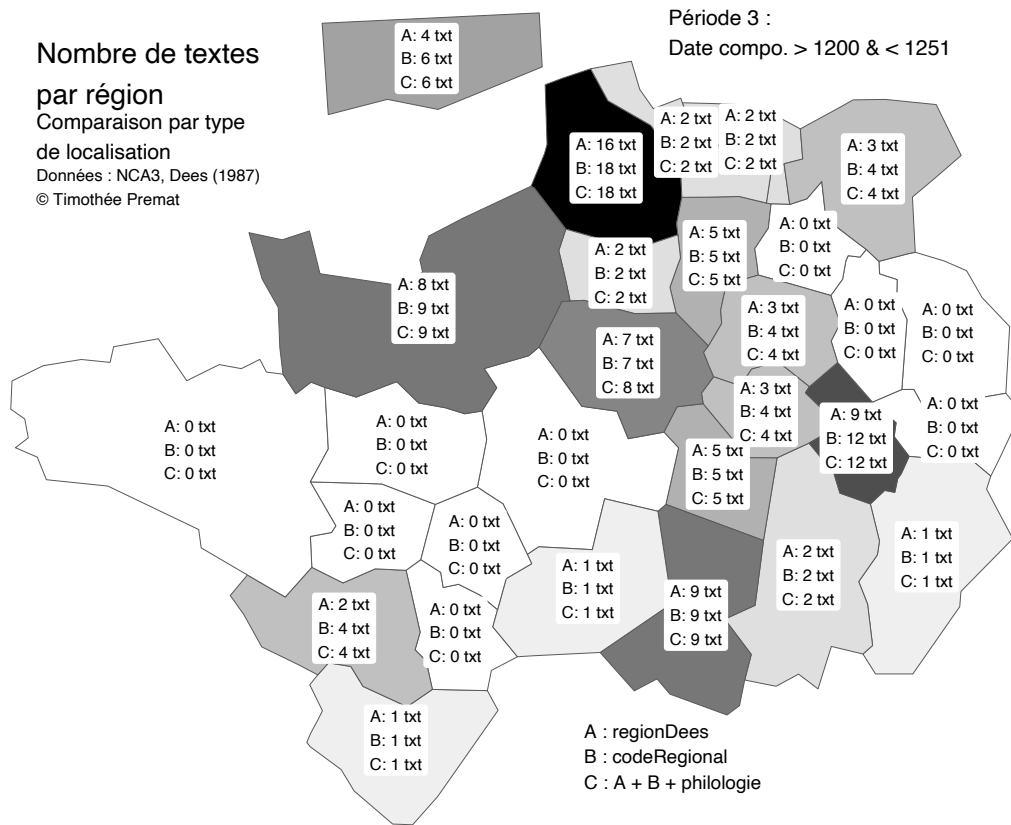


FIGURE 2.8 – Répartition des textes du NCA selon les localisations regionDees, complétées par codeRegional et par les informations philologiques.

a



b

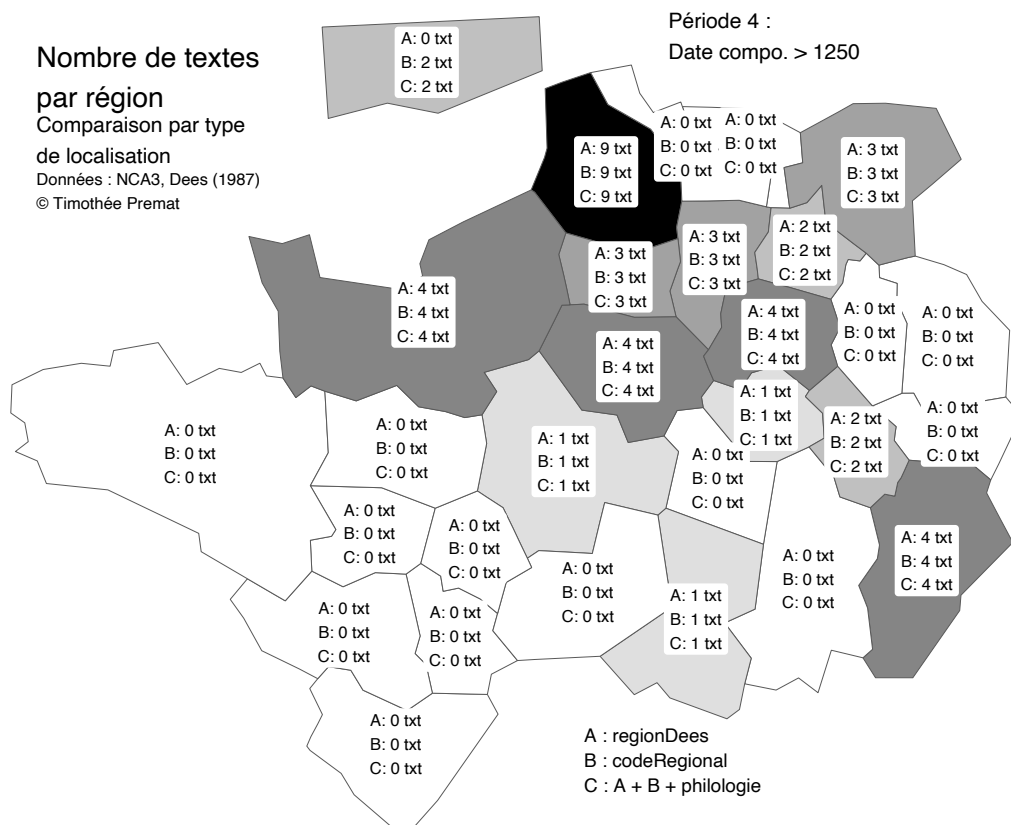


FIGURE 2.10 – Comparaison des nombres de textes par région selon les différents types de localisation, périodes 3 et 4.

marquante se joue en Angleterre (`regionDees` : 3 textes; `codeRegional` : 25 textes; localisation complétée : 35 textes), en Normandie (2 textes; 4 textes; 11 textes), en région Somme, Pas-de-Calais (10 textes; 12 textes; 15 textes), dans l'Aube (5 textes; 5 textes; 11 textes) ainsi que dans la région Vendée, Deux-Sèvres (3 textes; 8 textes; 8 textes). Ici encore, de nombreux textes disposent de la métadonnée `codeRegional`, mais sans disposer de la métadonnée `regionDees`. Parmi les textes qui disposent d'un coefficient dialectal, aucun ne dispose d'un coefficient inférieur à 70; la raison de leur exclusion est donc inconnue. L'usage de la complémentation des localisations de Dees (`regionDees` ou `codeRegional`) par les localisations philologiques ne parvient pas à combler de vides dans la carte : aucune région qui ne disposerait d'aucun texte chez Dees ne disposerait de textes grâce au recours, en sus, à la localisation philologique. En revanche, celle-ci permet d'améliorer sensiblement les taux de certaines régions; si cela semble être superflu pour l'Angleterre (qui dispose déjà de 25 textes avec les `codeRegionaux`) voire pour la région Somme, Pas-de-Calais (10 textes; 12 textes; 15 textes), c'est en revanche plus intéressant pour la Normandie (2 textes; 4 textes; 11 textes), la région parisienne (3 textes; 2 textes; 6 textes) et l'Aube (5 textes; 5 textes; 11 textes). Ainsi, dans le cas de ces trois dernières régions, le nombre de textes est plus que doublé par le recours aux localisations philologiques.

1201-1250. Dans la première moitié du ^{xiii}e s. (fig. 2.10 a), l'effet de nos manipulations sur les dates est peu visible : sur 99 textes, 85 disposent d'une `regionDees`. Sur les 14 textes restant, un seul ne dispose pas d'un `codeRegional` : `lechS`. Pour ce texte, la philologie propose la région parisienne (`compo. et ms. : frc.`), ce qui permet de faire passer cette région de 7 à 8 textes. De nouveau, les coefficients régionaux sont absents pour ces 14 textes sans `regionDees`. Le recours aux `codeRegionaux` permet donc de compléter le corpus par 13 textes, qui viennent renforcer la Haute-Marne (9 textes; 12 textes), l'Angleterre (4 textes; 6 textes), les régions Somme, Pas-de-Calais (16 textes; 18 textes) et Vendée, Deux-Sèvres (2 textes; 4 textes), ainsi que la Normandie (8 textes; 9 textes), la Wallonie (3 textes; 4 textes), la Marne et l'Aube (3 textes; 4 textes à chaque fois). Ici, c'est donc l'usage des `codeRegionaux` qui a le plus d'effet, en jouant notablement sur le taux de plusieurs régions mal représentées, les augmentant d'un tiers, ce qui n'est pas négligeable.

1251-1322. Enfin, la dernière période excède la seconde moitié du ^{xiii}e s. : nous la faisons aller jusqu'en 1322. En effet, après 1322, on ne trouve plus que trois textes dotés d'une date de composition, et ces trois textes, comme mentionné *supra*, se détachent du reste du corpus par une distance entre eux et le reste du corpus supérieure ou égale à 43 ans, distance sans égale dans le reste du corpus.³⁵ De 1251 à 1322, on trouve 43 textes, qui disposent tous d'une `regionDees`, à l'exception de `edmond` et `plainte`, pour lesquels le code régional et la philologie s'accordent sur une origine anglo-normande.

Ainsi, l'usage des localisations philologiques n'a d'effet décisif que dans la seconde moitié du ^{xiii}e s., où il permet de renforcer considérablement le poids de quelques régions. En aucun cas, une région n'est représentée que par un seul texte dont la localisation serait philologique, et, réciproquement, le recours aux localisations philologiques ne permet pas de

35: `loth` : `compo. 1365`; `songe14` : `compo. 1389`; `fierens` : `compo. 1489`.

comblent de véritables trous dans les cartes. L'effet étant limité, il nous semble que cet usage n'introduit pas de biais majeur dans notre démarche. Par ailleurs, si cet effet est limité, il se concentre sur la seconde moitié du ^{xix}^e s., et améliore donc sensiblement l'aspect quantitatif de cette période. Comme cet effet est par ailleurs intéressant, en ce qu'il permet de consolider certains résultats statistiques par une augmentation des témoins, il nous semble bienvenu. Dans ce qui suit, toutes les analyses seront menées sur la base des localisations complétées par la philologie, mais l'on prêtera toujours attention, dans la description de la diatopie, à ce que les textes dont la localisation ne vient pas de Dees ne conduisent pas à la génération de résultats aberrants. Par ailleurs, il importe, dans la lecture des cartes, de garder à l'esprit les deux points suivants : d'une part, en vertu des correspondances établies en table 2.4, l'Aube ne représente pas uniquement l'Aube mais a une valeur champenoise plus large que celle du département actuel. D'autre part, comme le suggèrent GLEÛGEN et GOUVERT (2007, p. 60-61), et comme c'est déjà en partie le cas chez Dees (par simple effet de la méthode dialectométrique), la région Parisienne représente autant une région géographique (en cohérence avec les chartes locales) qu'une région « sociolectale », et ce d'autant plus à mesure que les textes se font tardifs.

2.1.5.3 Métadonnées qualitatives

Comme évoqué *supra* et illustré en table 2.2, le corpus du NCA dispose de textes dont la qualité éditoriale varie. GLEÛGEN et GOUVERT (2007, p. 63) conseillent de n'utiliser que les textes de qualité CR1, CR2 et MS1 pour les études graphématiques ; on aurait pu construire un sous-corpus fiable, uniquement composé de textes à l'édition fidèle. Néanmoins, ce conseil ne s'applique que difficilement à une étude comme la nôtre, et ce pour plusieurs raisons. D'une part, les métadonnées qualitatives sont fragmentaires, avec 81 textes étiquetés MS et 17 textes CR, sans indice numérique de degré d'intervention. Retirer ces textes conduirait à retrancher le corpus de près d'un tiers de sa matière ; les intégrer reviendrait à introduire un biais, au regard de l'exclusion des textes MS3, CR2 et CR3.

Par ailleurs, et comme exposé dans les sous-sections précédentes, l'importance du nombre de textes est proportionnelle au nombre d'axes de la variation qu'une étude veut interroger. Ainsi, s'il serait éventuellement envisageable de se restreindre aux textes CR1, CR2 et MS1 pour une étude diachronique qui écraserait la diatopie, une simple étude diatopique se verrait lourdement handicapée par l'exclusion des textes MS, MS3, CR, CR2 et CR3. Dans une telle démarche, des régions dont les attestations sont déjà faibles au sein du corpus pourraient perdre jusqu'aux deux tiers de leurs textes, voire disparaître ; ainsi, comme illustré en table 2.5, la région Indre, Cher disparaîtrait tout simplement du corpus, tandis que les régions Bourgogne et Moselle, Meurthe-et-Moselle perdraient les deux tiers de leurs textes, et que les régions Yonne, Franche-Comté, Hainaut et Vendée, Deux-Sèvres perdraient entre la moitié et les deux tiers de leurs textes.

La situation devient encore plus critique lorsque l'on veut interroger l'intersection entre la diachronie et la diatopie ; de ce point de vue, au-delà même des quelques régions mentionnées en table 2.5, retirer presque un

TABLE 2.5 – NCA : régions qui perdraient la moitié ou plus de leurs textes par une restriction aux textes MS1, MS2 et CR1.

Région	nb. tot. textes	dont MS, CR, MS3, CR2, CR3
Indre, Cher	1	1 (1 MS)
Bourgogne	3	2 (2 CR)
Moselle, Meurthe-et-Moselle	3	2 (1 MS, 1 CR2)
Franche-Comté	5	3 (3 MS)
Yonne	5	3 (1 MS, 1 CR2, 1 CR3)
Vendée, Deux-Sèvres	12	7 (3 MS, 3 CR2, 1 CR3)
Hainaut	4	2 (1 MS, 1 MS3)

tiers du corpus n'est guère envisageable. Aussi, nous avons choisi de ne pas intégrer la variable qualitative dans notre chaîne de traitement du corpus. Néanmoins, l'information étant disponible dans les métadonnées à toutes les étapes du traitement, nous pourrions être amenés à remarquer l'impact de la qualité éditoriale sur les phénomènes observés.

2.1.6 En somme : le NCA ici utilisé

Le corpus NCA que nous exploitons est donc constitué de 290 textes, dont la date de composition va de 1100 (rolandox) à 1322 (contre); les textes postérieurs à 1322 ou sans date de composition sont systématiquement exclus par une ligne de code dans le traitement semi-automatisé des données. Sur ces 290 textes, 196 correspondent à des textes de l'*Atlas* de DEES (1987), et sont localisés par la métadonnée `regionDees`. Sur les textes qui n'appartiennent pas au corpus de l'*Atlas*, et pour lesquels la métadonnée `regionDees` manque, 57 sont néanmoins localisés par Dees, sous la forme du `codeRegional`, mais sans que leur coefficient de dialectalité ne soit indiqué. Enfin, pour 31 textes, c'est la philologie qui nous permet d'établir la localisation, tandis que 6 textes demeurent non localisés; ces textes apparaissent dans les résultats bruts et diachroniques mais sont omis lors de l'étude de la diatopie.

Dorénavant, chaque mention d'un texte issu de ce corpus s'accompagne d'une notice marginale qui reprend les informations essentielles, comme illustré ici pour `rolandox`. Cette note reprend toujours le format suivant :

- Identifiant;
- Auteur;
- Date de composition, dotée d'un astérisque lorsque la normalisation a effacé un *circa* ou une fourchette temporelle;
- Titre retenu par la tradition;
- Manuscrit en question, siglé selon la tradition philologique;
- Région attribuée par Dees;
- Lieu et date de composition selon le corpus philologique;
- Lieu et date de copie selon le corpus philologique;³⁶
- Qualité de l'édition;
- Entrée bibliographique de l'édition;
- Identifiant unique du texte au sein du DEAF, lorsque disponible.³⁷

rolandox : ANON. (1100*). *La Chanson de Roland*. Ms. O; Dees : Angleterre; compo. N.-O., ca 1100; ms. a.n., ca 1137; CR1 (SEGRE 1971); DEAF : Ro1S.

36: Pour les dates de composition et de copie, les fourchettes temporelles sont simplifiées en *ca*.

37: Nous ne reproduisons pas d'identifiant du DEAF lorsque le NCA propose un identifiant absent du DEAF.

L'ensemble des références des textes du NCA utilisés est indiqué en annexe, p. 596. Les références des éditions sont renvoyées en bibliographie générale.

2.2 BFM : la *Base de Français Médiéval*

Si le NCA représente le corpus principal sur lequel s'appuie cette thèse, nous recourons également à un second corpus numérique, lorsqu'il est ponctuellement question du très ancien français, ou d'analyse métrique automatique (chap. 12) : la BFM, ou *Base de Français Médiéval* (GUILLOT-BARBANCE, HEIDEN et LAVRENTIEV 2017). En effet, ce corpus contient les premiers monuments de la langue médiévale, et, contrairement au corpus de Dees, il n'a pas été nettoyé des accents que les éditeurs ajoutent sur les <e> finaux lorsque ceux-ci représentent un /e/ tonique et non un schwa, et est systématiquement découpé en vers, ces deux points étant importants pour notre programme d'analyse métrique. Dans la mesure où nous n'utilisons pas l'ensemble du corpus de la BFM, il ne s'agit pas ici de documenter exhaustivement ce corpus, mais uniquement d'en exposer les caractéristiques principales et les propriétés pertinentes pour notre travail. Par ailleurs, l'analyse des textes médiévaux avec le PAM (Programme d'Analyse Métrique) n'a fait que confirmer le fait que les élisions métriques étaient celles que supposait la théorie métrique, et n'a pas permis de dégager de tendance fiable. Aussi, les citations du corpus de la BFM dans ce travail sont ponctuelles, même si un travail d'ampleur a été effectué avec le PAM sur ce corpus.

Le corpus de la BFM a été fondé en 1989 par Christiane Marcello-Nizia et est aujourd'hui dirigé par Céline Guillot-Barbance. Nous l'utilisons dans sa version 2019 ; sauf mention du contraire, toute référence au corpus de la BFM dans ce texte sous-entend sa version 2019. Dans cette version, le corpus contient 170 textes intégraux, du ix^e s. au xv^e s., pour un ensemble de près de 4 700 000 occurrences-mots.³⁸ Peu de textes sont des éditions nativement numériques ; la plupart ont été établis à partir d'éditions imprimées existantes, parfois anciennes, sans modification de cette édition, et sans vérification du protocole éditorial adopté.

Le corpus de la BFM est consultable et interrogeable en ligne, sur le portail de la BFM, par le biais de l'interface web du logiciel TXM (HEIDEN, MAGUÉ et PINCEMIN 2010). Le corpus est également accessible en téléchargement sur demande auprès de l'équipe de la BFM. L'ensemble des textes est au format .xml et répond aux préconisations du consortium *Text Encoding Initiative* dans leur version 5 (TEI-P5).³⁹ Dans la version accessible en ligne, seule une partie du corpus est annotée en parties du discours et en lemmes. Cette annotation a été produite par le logiciel TreeTagger, puis a été vérifiée humainement pour réduire la marge d'erreur de l'annotation automatique. Dans la version téléchargeable sur demande, le corpus peut avoir deux formes : soit les fichiers téléchargés sont les fichiers .xml de chaque texte, auquel cas l'annotation est exactement la même que sur l'interface en ligne, soit les fichiers téléchargés sont des fichiers binaires à destination de TXM, auquel cas on peut obtenir une lemmatisation et une annotation en partie du discours de tous les textes. La version avec laquelle nous travaillons est la version binaire entièrement annotée.

GUILLOT-BARBANCE et al. (2017). « Base de français médiéval : une base de référence de sources médiévales ouverte et libre au service de la communauté scientifique ».

38: <http://bfm.ens-lyon.fr/>

HEIDEN et al. (2010). « TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement ».

39: <https://tei-c.org/>

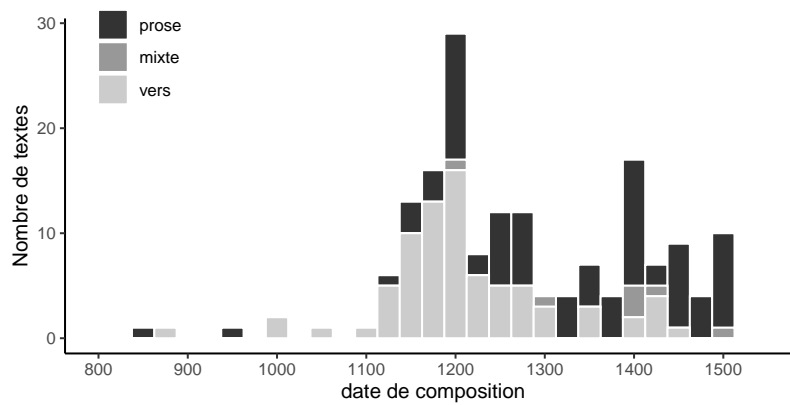


FIGURE 2.11 – Répartition des textes de la BFM selon la date de composition normalisée.

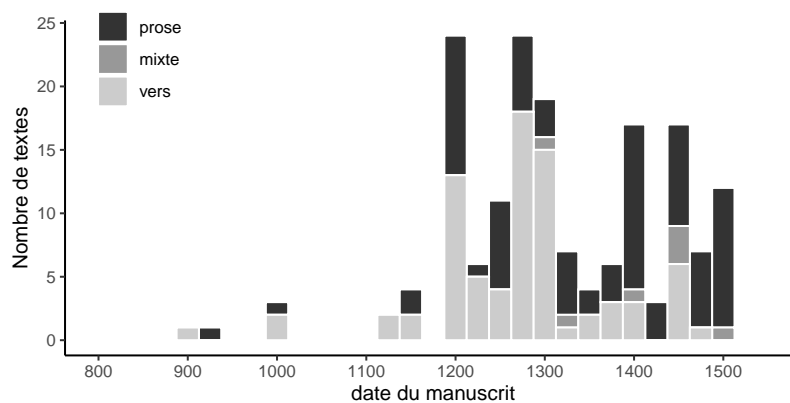


FIGURE 2.12 – Répartition des textes de la BFM selon la date du manuscrit normalisée.

2.2.0.1 Composition du corpus

Sur les 170 textes intégraux de la BFM, 84 sont en prose, 78 en vers et 8 sont mixtes. Au sein de ce corpus, 34 textes ont fait l'objet d'une vérification manuelle intégrale de l'annotation morphosyntaxique ; 11 sont en prose, 19 en vers et 4 sont mixtes. Par ailleurs, 5 textes ont fait l'objet d'une vérification partielle de l'annotation morphosyntaxique ; 3 sont en prose et 2 en vers. Enfin, 16 textes ont fait l'objet d'une vérification manuelle de la lemmatisation : 5 sont en prose, 10 sont en vers et 1 est mixte, tandis que deux textes en prose ont fait l'objet d'une vérification partielle des lemmes. Au total, 46% des textes sont intégralement versifiés.

Les textes de la BFM vont du ix^e siècle (*Serments de Strasbourg*, datés de 842) à la fin du xv^e s. (*Mémoires de Philippe de Commines*, compo. ca. 1490-1505). Les figures 2.11 et 2.12 représentent la distribution des textes au sein du corpus de la BFM, après normalisation des dates.⁴⁰ L'écart entre date de composition et date de copie va de 0 à 237 ans ; l'écart moyen est de 44 ans. Le peu de textes avant 1100 reflète le peu d'attestations de la littérature vernaculaire avant le xii^e s. et n'est pas imputable au corpus. La domination des textes en prose à partir de 1200 (date de composition) représente également la dynamique à l'œuvre dans la production littéraire française et n'est pas un biais du corpus.

40: Cette normalisation est du fait de l'auteur de ces lignes ; elle s'est effectuée selon le principe suivant : médiane en cas de fourchette de dates, début du siècle $n = n + 10$ ans, fin du siècle $n = n - 10$ ans, etc.

La BFM assigne une étiquette dialectale à chaque texte. Pour 40 textes, le dialecte est qualifié de « non défini ». Les dialectes les plus représentés au sein du corpus sont l'anglo-normand, le picard, le champenois et le normand, ce qui suit encore une fois plus ou moins bien l'état de la documentation historique, ces régions étant des centres importants pour la copie des textes. Dans certains cas, la BFM propose des étiquettes mixtes (« picard / wallon ») ou plus larges qu'un seul dialecte (« Ouest », « Nord-Ouest », « Nord »). Deux étiquettes mixtes, « franco-italien » et « franco-occitan », ne s'appliquent qu'à deux textes chacune, respectivement le *Roman de Renart*, branches x (renart10) et x1 (renart11), et la *Passion de Clermont* (passion) ainsi que la *Vie de saint Léger* (slethgier). Ces deux derniers textes, appartenant au même manuscrit auvergnat, ont fait l'objet de vives controverses quant au statut de leur langue *mixte* (cf. p.ex. PARIS 1872, p. 275-286, non pour ses conclusions mais pour l'exposé du débat). La table 2.6 donne le nombre de textes par dialecte ; la carte en fig. 2.13⁴¹ donne en sus la répartition siècle par siècle de ces textes. La liste des textes cités et de leurs métadonnées pertinentes est donnée en annexe, p. 608.

(PARIS 1872). « La *Vie de saint Léger*, texte revu sur le ms. de Clermont-Ferrand ».

41: Un seul texte, prov, une collection de proverbes, n'est pas affiché sur la carte, en raison de l'absence de date précise. Il n'a pas de dialecte identifié.

TABLE 2.6 – BFM : nombre de textes par dialecte.

dialecte	nb.	dialecte	nb.
non défini	39	wallon	4
anglo-normand	24	picard / wallon	3
picard	21	bourguignon	2
champenois	13	franco-italien	2
Ouest	12	franco-occitan	2
normand	11	aube	1
orléanais	8	flandres	1
parisien	8	Nord	1
lorrain	6	haineau	1
Nord-Ouest	5	liégeois	1
poitevin-saintongeais	4		

2.2.0.2 Structure de l'annotation

Les textes du corpus de la BFM sont formatés en langage XML dans le respect des normes de la TEI-P5. Chaque fichier, après un préambule contenant l'ensemble des méta-données, présente le texte au sein de l'environnement de corps de texte <body>. Lorsque le texte est annoté en morphosyntaxe et en lemmes, la présentation est la suivante : chaque mot est enserré dans un environnement ouvert et fermé par les balises <w> et </w>. ⁴² Dans la balise d'ouverture sont définis la catégorie morphosyntaxique du mot (type), le dictionnaire d'où provient le lemme (ana⁴³) et le lemme (lemma). Les retours à la ligne sont marqués par une balise <lb>⁴⁴ en début de ligne ; dans les textes versifiés, c'est cette balise qui découpe et numérote les vers. L'environnement <s>⁴⁵ permet également le découpage et la numérotation des phrases. Cette présentation est reproduite en fig. 2.14 pour les deux premiers vers de la *Chanson de Roland* : *Carles li reis, nostre emperere magnes // Set anz tuz pleins ad estet en Espagne*. Le mot annoté est donc placé à l'extrémité de chaque ligne.

42: w pour *word*.

43: ana pour *analyse*, en référence au dictionnaire utilisé pour la dite analyse.

44: lb pour *line break*.

45: s pour *sentence*.

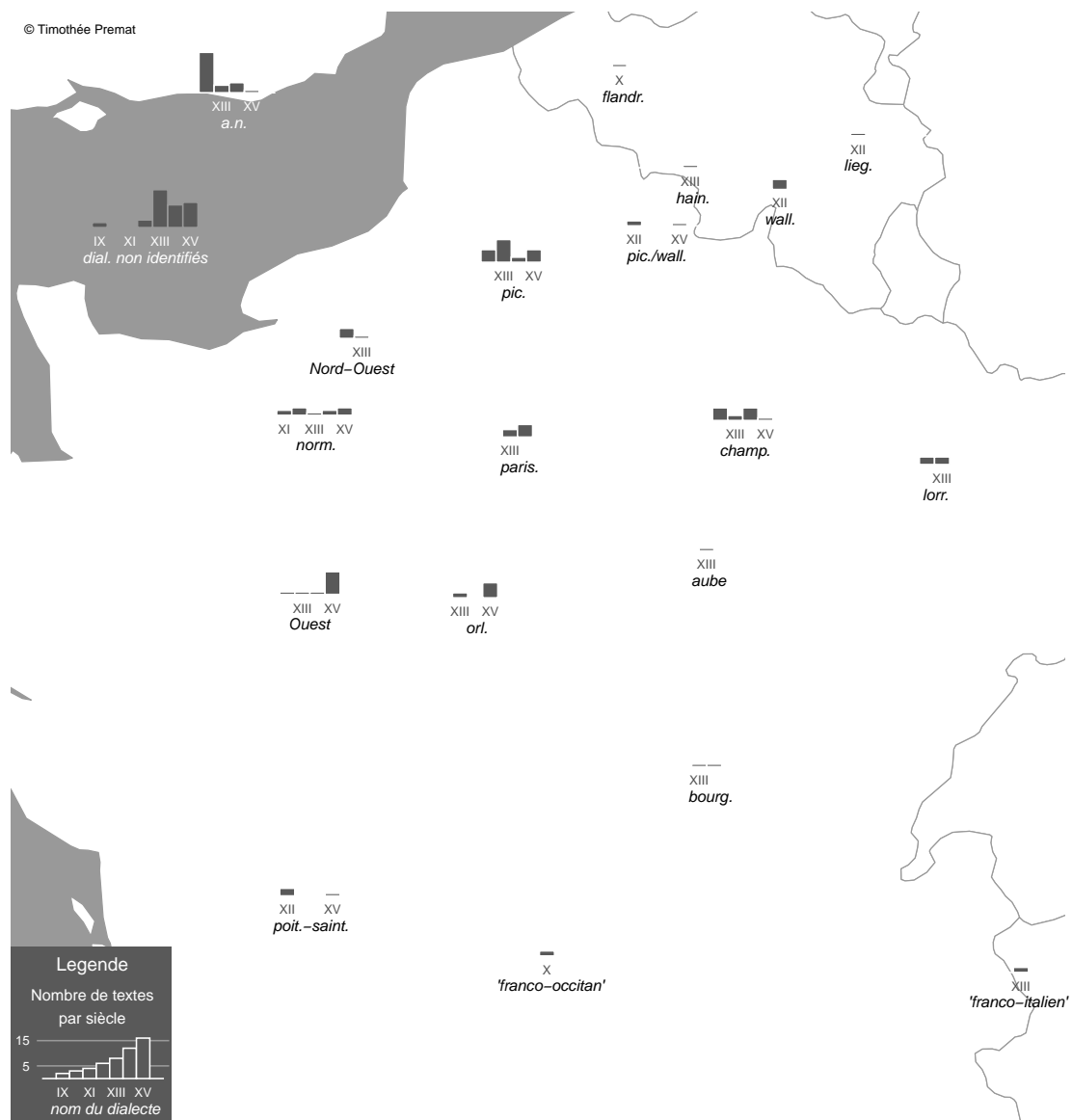


FIGURE 2.13 – Distribution des textes de la BFM par dialectes et par siècle (date de composition normalisée).

Sur la version publique des textes de la BFM, telle qu'accessible via le portail internet, le niveau d'annotation et le format fluctuent. L'annotation de roland en fig. 2.14 est celle d'un texte dont la morphosyntaxe et la lemmatisation ont été vérifiées; seules les annotations vérifiées sont accessibles en ligne. Lorsqu'un texte n'a pas fait l'objet d'une annotation vérifiée, les mots ne sont pas insérés dans des balises et sont simplement séparés par des espaces. Les retours à la ligne marquent, dans les textes versifiés, le début des vers – ce qui sera important pour l'utilisation du programme d'analyse métrique PAM. En fig. 2.15 sont représentés les quatre premiers vers de la *Chronique des Ducs de Normandie* de Benoît de Sainte Maure (BenDuc1b), qui représente le formatage des fichiers sans annotation.

```

<lb n="1"/>
<s n="1">
<w type="NOMpro" ana="#lemma_src_DMF" lemma="Charles" xml:id="w_58_00001">Carles</w>
<w type="DETdef" ana="#lemma_src_DMF" lemma="le" xml:id="w_58_00002">li</w>
<w type="NOMcom" ana="#lemma_src_DMF" lemma="roi1" xml:id="w_58_00003">reis</w>
<w type="PONfbl" ana="#lemma_src_DMF" lemma="," xml:id="w_58_00004">,</w>
<w type="DETpos" ana="#lemma_src_DMF" lemma="notre" xml:id="w_58_00005">nostre</w>
<w type="NOMcom" ana="#lemma_src_DMF" lemma="empereur" xml:id="w_58_00006">emperere</w>
<w type="ADJqua" ana="#lemma_src_DMF" lemma="maigne" xml:id="w_58_00007">magnes</w>
<w type="PONfbl" ana="#lemma_src_DMF" lemma="," xml:id="w_58_00008">,</w>
<lb n="2"/>
<w type="DETCar" ana="#lemma_src_DMF" lemma="sept" xml:id="w_58_00009">Set</w>
<w type="NOMcom" ana="#lemma_src_DMF" lemma="an" xml:id="w_58_00010">anz</w>
<w type="ADVgen" ana="#lemma_src_DMF" lemma="tout2" xml:id="w_58_00011">tuz</w>
<w type="ADJqua" ana="#lemma_src_DMF" lemma="plein" xml:id="w_58_00012">pleins</w>
<w type="VERcjpg" ana="#lemma_src_DMF" lemma="avoir1" xml:id="w_58_00013">ad</w>
<w type="VERppe" ana="#lemma_src_DMF" lemma="être1" xml:id="w_58_00014">estet</w>
<w type="PRE" ana="#lemma_src_DMF" lemma="en1" xml:id="w_58_00015">en</w>
<w type="NOMpro" ana="#lemma_src_DMF" lemma="Espagne" xml:id="w_58_00016">Espagne</w>
<w type="PONfbl" ana="#lemma_src_DMF" lemma=":" xml:id="w_58_00017">:</w>

```

FIGURE 2.14 – *Chanson de Roland*, v. 1-2, dans le format du fichier roland.xml de la BFM.

```

<lb n="10541"/>Quant li buens dux fu trepassez,
<lb n="10542"/>Qui tant fu au siecle onnoez,
<lb n="10543"/>Si furent tuit mandé par non
<lb n="10544"/>Ensemble Normant e Breton.

```

FIGURE 2.15 – *Chronique des Ducs de Normandie* de Benoît de Sainte Maure, v. 1-4, dans le format du fichier BenDuc1b.xml de la BFM.

2.2.1 Comparaison entre NCA et BFM

L'annotation du premier vers de la *Chanson de Roland* dans le NCA est reproduite en fig. 2.16, pour permettre une comparaison de la structure de l'annotation, vis-à-vis de l'annotation du même vers dans la BFM, en fig. 2.14. Il faut principalement constater la plus grande richesse de l'annotation du NCA. Cette richesse tient pour partie à la présence des étiquettes de Dees (deespos) et de leur remaniement par le NCA (pos), en concurrence avec les étiquettes de TreeTagger (taggerpos et lemma). Mais, au-delà du simple nombre d'attributs, cette richesse est aussi qualitative : si les étiquettes de TreeTagger sont équivalentes entre les deux corpus (BFM NOMpro = NCA NPR, etc.), les étiquettes pos du NCA ajoutent des informations relatives au genre, au nombre, au cas ou à la conjugaison du mot étiqueté. Ainsi, dans *Carles li reis*, le déterminant *li* est étiqueté DETdef et DET_def par TreeTagger dans la BFM et le NCA, mais, dans le NCA, l'attribut pos issu des étiquettes remaniées de Dees précise que le déterminant défini est au cas sujet masculin singulier (DET_def_suj_masc_sg). Si cet étiquetage automatique n'est pas sans erreurs, celles-ci demeurent relativement marginales : lors du dépouillement et de la vérification des résultats, peu d'occurrences ont dû être retranchées à cause de problèmes analytiques dans cet attribut pos. La présence de cette information flexionnelle est évidemment importante pour notre

```

1 <s n="nca:355312_rolandox:1">
  <w id="w_rolandox_s355312_1">
    <form>
      <form>carles</form>
      <ana type="#taggerpos">NPR</ana>
      <ana type="#pos">NPR_suj_masc_sg</ana>
      <ana type="#lemma">charle</ana>
      <ana type="#deespos">011</ana>
    </form>
    <ana type="#type">w</ana>
    <ana type="#n">6</ana>
  </w>
3 <w id="w_rolandox_s355312_2">
  <form>
    <form>li</form>
    <ana resp="#src" type="#taggerpos">DET_def</ana>
    <ana resp="#src" type="#pos">DET_def_suj_masc_sg</ana>
    <ana resp="#src" type="#lemma">le</ana>
    <ana resp="#src" type="#deespos">101</ana>
  </form>
  <ana resp="none" type="#type">w</ana>
  <ana resp="none" type="#n">7</ana>
</w>
4 <w id="w_rolandox_s355312_3">
  <form>
    <form>reis</form>
    <ana resp="#src" type="#taggerpos">NOM</ana>
    <ana resp="#src" type="#pos">NOM_suj_masc_sg</ana>
    <ana resp="#src" type="#lemma">roi</ana>
    <ana resp="#src" type="#deespos">001</ana>
  </form>
  <ana resp="none" type="#type">w</ana>
  <ana resp="none" type="#n">8</ana>
</w>
5 <w id="w_rolandox_s355312_4">
  <form>
    <form>nostre</form>
    <ana resp="#src" type="#taggerpos">DET_poss</ana>
    <ana resp="#src" type="#pos">DET_poss_suj_masc_sg</ana>
    <ana resp="#src" type="#lemma">nostre</ana>
    <ana resp="#src" type="#deespos">161</ana>
  </form>
  <ana resp="none" type="#type">w</ana>
  <ana resp="none" type="#n">9</ana>
</w>
6 <w id="w_rolandox_s355312_5">
  <form>
    <form>empere</form>
    <ana resp="#src" type="#taggerpos">NOM</ana>
    <ana resp="#src" type="#pos">NOM_suj_masc_sg</ana>
    <ana resp="#src" type="#lemma">empire</ana>
    <ana resp="#src" type="#deespos">001</ana>
  </form>
  <ana resp="none" type="#type">w</ana>
  <ana resp="none" type="#n">10</ana>
</w>
7 <w id="w_rolandox_s355312_6">
  <form>
    <form>magnes</form>
    <ana resp="#src" type="#taggerpos">ADJ</ana>
    <ana resp="#src" type="#pos">ADJ_suj_masc_sg</ana>
    <ana resp="#src" type="#lemma">maigne</ana>
    <ana resp="#src" type="#deespos">021</ana>
  </form>
  <ana resp="none" type="#type">w</ana>
  <ana resp="none" type="#n">11</ana>
</w>
8 </s>

```

FIGURE 2.16 – *Chanson de Roland*, v. 2, dans le format du fichier rolandox.xml du NCA. Le fichier a ici été simplifié (les éléments dus à l'importation dans TXM sont non pertinents pour cette discussion), et mis en page pour mettre en évidence la structure arborescente du XML; la version fidèle au XML retravaillé par TXM est rejetée en annexes, en fig. E.1 (p. 610).

étude, dans la mesure où elle permet, par exemple, de distinguer les adjectifs masculins et féminins. Sans une telle information, toute requête portant sur les adjectifs serait aveugle à leur flexion en genre, et il serait impossible de repérer les adjectifs qui ont une variation $-ə/\emptyset$ non liée au genre. De même, pour interroger les substantifs et adjectifs masculins de la première déclinaison, il faut avoir accès à leur cas : dans la mesure où les cas sujet singulier et cas régime pluriel se terminent par un $-s$, tous les substantifs et adjectifs de cette déclinaison qui se terminent par une séquence $/əs/$ présentent une alternance finale entre $\langle e \rangle$ pour schwa et $\langle es \rangle$, laquelle n'est pas pertinente pour notre étude. La granularité de l'annotation du NCA est donc un précieux atout pour notre étude.

En revanche, là où la BFM l'emporte, pour notre étude, sur le NCA, c'est sur la présence de diacritiques éditoriaux, de marques délimitant les vers et de textes en très ancien français. Ainsi, le participe *mandé* en fig. 2.15 serait devenu *mande* selon le traitement de Dees, restituant par là la probable forme du manuscrit, mais la rendant homographe des formes verbales. De même, les trémas éditoriaux sont absents du NCA : l'infinif *oir* dans la BFM correspond à *oir* dans le NCA.⁴⁶ Or le programme informatique que nous utilisons pour le traitement automatique de la métrique, le PAM, a besoin de ces distinctions, pour calculer si la syllabe est accentuable ($\langle e \rangle$ pour $/e/$) ou élidable ($\langle e \rangle$ pour $/ə/$), et pour calculer s'il doit considérer une suite de graphèmes vocaliques comme une di-ou triptongue ou comme un hiatus (*o.ïr*) ; ce programme ne peut donc pas être déployé sur le NCA sans un imposant processus de retravail du corpus en amont. Par ailleurs, toujours sur la question des textes versifiés, le découpage des vers est certes polymorphe dans la BFM, mais il est toujours présent : les vers sont représentés soit par des balises $\langle lb \rangle$, soit par des retours à la ligne, alors que le découpage des vers fait défaut dans une partie des textes versifiés du NCA. Enfin, contrairement au NCA dont le plus ancien texte est la *Chanson de Roland* (compo. ca 1100), la BFM contient six textes antérieurs à 1100 : les *Serments de Strasbourg* (*strasbBfm*), la *Séquence de Sainte Eulalie* (*eulaliBfm*), le *Sermon sur Jonas* (*jonas*), la *Passion de Clermont* (*passion*), la *Vie de saint Léger* (*sléthgier*) et la *Vie de saint Alexis* (*AlexisRaM*). De plus, pour trois de ces textes, les éditions sont diplomatiques et nativement numériques : l'équipe de la BFM a édité elle-même les *Serments de Strasbourg*, la *Séquence de sainte Eulalie* et la *Vie de saint Alexis*, produisant même pour cette dernière une édition multi-niveaux qui rend disponible, en regard de l'édition normalisée, une édition diplomatique fine. En revanche, nous ne recourons pas à la *Passion de Clermont* et à la *Vie de Saint Léger*, dans la mesure où, comme exposé *supra*, leur langue présente un mélange de traits occitans et oiliques qui ne permet aucune analyse assurée en ce qui concerne la phonologie de l'oïl archaïque.

Ainsi, le recours à la BFM nous permettra d'étudier la période archaïque de l'ancien français, ainsi que la métrique des textes versifiés, par le biais du PAM. Si des interrogations lexicales peuvent aussi, ponctuellement, référer au corpus de la BFM, la quatrième partie de cette thèse, elle, portera uniquement sur le corpus du NCA.

46: *Aucassin et Nicolette*, v. 1 : *qui vauroit bons vers oïr* (BFM, *aucassin*) ~ *qui vauroit bons vers oir* (NCA, *aucchants*).

aucassin : ANON. (1208*). *Aucassin et Nicolette*. Compo. dern. q. XII^e s. ou 1^{ère} m. XIII^e s., ms. fin XIII^e s., dial. pic. (ROQUES 1929a).

aucchants : ANON. (1225*). *Aucassin et Nicolette*. Dees : Pas-de-Calais ; compo. pic., ca 1225 ; ms. pic., ca 1290 ; MS1 (ROQUES 1955) ; DEAF : AucR³ ; ne contient que les laisses chantées.

strasbBfm : ANON. (842*). *Serments de Strasbourg*. Compo. ca 842, ms. ca 1000, (LAVRENTIEV, RAINSFORD et GUILLOT-BARBANCE 2014).

eulaliBfm : ANON. (881*). *Séquence de sainte Eulalie*. Compo. peu après 881, ms. ca 900, (LAVRENTIEV, GUILLOT-BARBANCE et RAINSFORD 2018).

jonas : ANON. (ca 945*). *Sermon sur Jonas*. Compo. prob. entre 938 et 952, ms. 2^e q. X^e s., dial. fla. (POERCK 1956).

passion : ANON. (1000*). *Passion de Jésus-Christ ou Passion de Clermont*. Compo. et ms. ca 1000, dial. fr.-oc. (AVALLE 1962).

sléthgier : ANON. (1000). *Vie de saint Léger*. Compo. ca 1000, ms. ca 1000, dial. fr.-oc. (AVALLE 1967).

AlexisRaM : ANON. (1050*). *Vie de saint Alexis*. Compo. ca 1050, ms. ca 1120, dial. norm. (RAINSFORD et MARCHELLO-NIZIA 2018).

2.3 Sources musicales : philologie musicale des XII^e et XIII^e s.

En marge des analyses informatiques ou manuelles que nous faisons des corpus NCA et BFM, des éléments de musication nourrissent également notre analyse – dans ce travail, *musication* désigne l’association d’un texte, toujours versifié à l’époque médiévale, avec une musique. Comme la métrique, l’analyse de la musication est essentiellement mobilisée lorsque nous interrogeons les élisions non transcrites par la graphie (chap. 13). Pour l’analyse de la musication, nous nous appuyons toujours sur les sources manuscrites. Si la production de corpus au format XML-MEI (équivalent musical du XML-TEI) semble se développer en musique ancienne, nous n’avons pas connaissance de corpus de ce type qui soient aboutis et qui concernent la musique médiévale de langue d’oïl. Au demeurant, le recours à une édition est encore plus problématique pour les sources musicales que pour les sources textuelles : au-delà des simples choix éditoriaux et interprétatifs de l’éditeur, la notation moderne est impropre à rendre compte d’un certain nombre d’éléments transmis par les notations médiévales. Aussi, toutes nos analyses se sont toujours faites soit directement sur le manuscrit, soit à l’aide d’une édition, mais avec une vérification scrupuleuse de la leçon du manuscrit pour les éléments analysés.

Un corpus musical principal est mobilisé dans cette thèse : le corpus des trouvères, vaste ensemble de la fin du XII^e et du XIII^e s. qui contient les premières musiques vocales de langue d’oïl dont la musique nous soit parvenue. S’y ajoutent, de manière plus ponctuelle, des références à des pièces isolées, issues d’*Aucassin et Nicolette* et du *Gieus de Robin et Marion* (par ailleurs présents dans le NCA : *aucchants* et *robin*). Ces corpus ont déjà fait l’objet d’études précédentes de notre part (PREMAT et CHOUVION 2021; PREMAT, CHOUVION et VERNER 2021; CHOUVION, PREMAT et VERNER 2021; MORÉE et PREMAT 2021), comme corpus principal pour la première étude, et comme corpus de comparaison dans les autres études. Si les chansons de trouvères constituent les premières pièces musicales dont la musique nous soit parvenue, il importe de signaler que la musique en langue d’oïl précède évidemment les trouvères ; il est ainsi à peu près certain que les chansons de gestes, forme archaïque par excellence de la littérature d’oïl, étaient récitées sous une forme plus ou moins musicale, mais leur musique n’a pas été transcrite.⁴⁷

Les sources musicales ayant une dimension philologique spécifique, cette section détaille les différents corpus utilisés, les caractéristiques pertinentes des manuscrits utilisés et avance quelques notions musicologiques utiles à l’analyse. Ainsi, nous détaillons dans la section suivante les sources utilisées et les limites dans l’interprétation que nous nous imposons pour garantir la qualité des données analysées.

2.3.1 Chansons de trouvères, *Aucassin et Nicolette* et *Li Gieus de Robin et de Marion*

Nous avons étudié, avec une musicologue, 159 chansons de trouvères (PREMAT et CHOUVION 2021). Ces chansons sont issues de neuf manuscrits, appelés par la tradition a, i, K, M, N, O, P, U et X, disponibles en ligne, sur

aucchants : ANON. (1225*). *Aucassin et Nicolette*. Dees : Pas-de-Calais ; compo. pic., ca 1225 ; ms. pic., ca 1290 ; MS1 (ROQUES 1955) ; DEAF : AucR³ ; ne contient que les laisses chantées.

robin : ADAM DE LA HALLE (1225*). *Li Gieus de Robin et Marion*. Ms. P ; Dees : Somme ; compo. art., ca 1285 ; ms. Arras, 1295 ; MS (ADAM DE LA HALLE 1980).

PREMAT et CHOUVION (2021). « Adaptations de la mélodie aux variations de longueur du vers aux XII^e et XIII^e s. »

PREMAT, CHOUVION et VERNER (2021). « Repetitions, rests insertion and Schwa in 16th century French Polyphony : An Emergence Sub-Grammar in Fresneau’s Songs ? »

CHOUVION, PREMAT et VERNER (2021). « La musication dans les chansons d’Henry Fresneau, Grammaire d’association, parophonologie et forme polyphonique au XVI^e siècle ».

MORÉE et PREMAT (2021). « Multilingualism in Music: A Multidisciplinary Analysis of the Codex Reina and its Franco-Dutch repertoire (c. 1400) ».

47: Trois témoins de la musicalité des chansons de gestes demeurent sous forme notée, mais ils sont tardifs et parodiques : *Aucassin et Nicolette* est considéré comme une parodie épique, sur la forme comme sur le fond ; le *Gieus de Robin et de Marion* contient une parodie épique, l’un des personnages déclarant *Je sai trop bien canter de geste / Me volés vous oïr canter* suivi d’une ligne mélodie correspondant à un vers grossier, avant de se faire moquer par son auditoire (fol. 48v, v. 726-732 ; leçon de BETTENS (ADAM DE LA HALLE 2002 [1998])) ; enfin, la courte parodie épique appelée *Bataille d’Annezin* comprend, à son terme, une brève phrase musicale notée, reproduite en 2.17, plus ornée que les autres exemples mais respectant certaines caractéristiques musicales du genre (FÉRAUD 2016).



FIGURE 2.17 – Phrase musicale de la *Bataille d’Annezin*, British Library Royal Collection BL Royal 20 A XVII, fol. 177r. Url : http://www.bl.uk/manuscripts/Viewer.aspx?ref=royal_ms_20_a_xvii_f177r.

le portail de la Bibliothèque Vaticane pour a, sur le portail Gallica de la BNF pour les autres. Ces chansons ont fait l'objet d'une édition musicale et textuelle par ROSENBERG, SWITTEN et LE VOT (1998) et/ou ROSENBERG, TISCHLER et GROSSEL (1995). Bien qu'ayant un éditeur en commun, ces deux ouvrages diffèrent : tandis que le premier n'édite qu'un petit nombre de chansons, sans notation rythmique la plupart du temps, et avec un appareil analytique et philologique fouillé, le second édite un grand nombre de chansons, en notation rythmique moderne, et avec seulement quelques notes de fin d'ouvrage. Les références des ms. sont reproduites en annexes, p. 609.

Hormis U, qui date selon toute vraisemblance de la première moitié du XIII^e s., tous les autres manuscrits sont tardifs : datant de la fin du XIII^e s. voire du XIV^e s., les autres témoins arrivent à la fin de la période où se développe l'art des trouvères (PARKER 1978, p. 3). Si U est isolé par sa précocité et sa notation musicale, il l'est aussi par son relatif isolement philologique, tandis que les manuscrits tardifs peuvent faire l'objet d'une analyse phylogénétique, soit simplement au niveau du texte (SCHWAN 1886), ou au niveau du texte et de la musique (PARKER 1978). De ce point de vue, les ms. K, N, P et X présentent un fort degré de ressemblance et sont apparentés, que ce soit par Schwan ou par Parker. Sur le fondement d'analyses de la mouvance musicale, PARKER (1978, p. 182-184) propose d'aller plus loin, et de voir P et X comme enfants, dans le stemma, de K, celui-ci étant placé au même niveau que N. Nos autres manuscrits sont relativement indépendants.

2.3.1.1 Les témoins des trouvères

Nous détaillons ici le strict nécessaire des informations philologiques relatives aux manuscrits des trouvères utilisés dans nos études.

Le ms. U. Le ms. siglé U par la tradition correspond au ms. BNF fr. 20050, également appelé *Chansonnier de Saint-Germain-des-Prés*. Si certaines pièces sont sans musique notée (les portées ont été dessinées mais n'ont pas été remplies), il est d'une copie soignée. Sa datation habituelle oscille entre 1240 et 1268, ou parfois peu avant (GROSSEL 2016). Ce témoin est donc extrêmement précieux, en ce qu'il représente la plus ancienne attestation des chansons de langue d'oïl. Au-delà de sa précocité chronologique, U se distingue aussi par sa notation : il est écrit en notation messine (graphie musicale originaire de Metz), à une époque, le XIII^e s., où les notes carrées, liées aux contraintes du placement des notes sur la portée et à l'évolution technique des plumes, est « définitivement adoptée » comme norme de « notation unique », au détriment des graphies musicales régionales pré-existantes (DENNERY 1982, p. 101). Ainsi, le signe de base du ms. U, la *virga* ✦, est encore relativement proche de la *virga* paléofranque / des premières notations, qui n'est autre qu'un accent aigu, tandis que nos autres témoins sont tous en notes carrées, où la *virga* n'a plus de verge que le nom : ♣. Le ms. U est donc doublement archaïque : d'une part par sa datation et l'état du répertoire qu'il contient, et d'autre part par sa notation déjà dépassée.

ROSENBERG et al. (1998). *Songs of the troubadours and trouvères: an anthology of poems and melodies*.

ROSENBERG et al. (1995). *Chansons des trouvères, chanter m'estuet*.

PARKER (1978). « À propos de la tradition manuscrite des chansons de trouvères ».

SCHWAN (1886). *Die altfranzösischen Liederschriften : ihr Verhältnis, ihre Entstehung und ihre Bestimmung*.

U : *Chansonnier français, dit de Saint-Germain-des-Prés* (BNF fr 20050). URL : <http://gallica.bnf.fr/ark:/12148/btv1b60009580>.

GROSSEL (2016). « Madeleine Tyssens, *Le Chansonnier français U* ».

DENNERY (1982). « Les notations musicales au moyen âge ».

N : *Français 845* (BNF fr 845). URL : <http://gallica.bnf.fr/ark:/12148/btv1b6000955r>.

P : *Français 847* (BNF fr 847). URL : <http://gallica.bnf.fr/ark:/12148/btv1b8454673n>.

X : *Chansons françaises, avec musique notée (Chansonnier Clairambault)* (BNF fr 1050). URL : <http://gallica.bnf.fr/ark:/12148/btv1b530003205>.

K : *Recueil de chansons du XIII^e siècle, avec musique notée* (BNF fr 5198). URL : <http://gallica.bnf.fr/ark:/12148/btv1b550063912>.

M : *Français 844* (BNF fr 844). URL : <https://gallica.bnf.fr/ark:/12148/btv1b84192440>.

Les ms. M, N, P, X et K. Les ms. siglés N, P, X et K par la tradition forment une famille cohérente (PARKER 1978), tant par les pièces qui y sont reproduites, par leur présentation, que par la relative conformité de leurs leçons. Le ms. M n'est que plus faiblement rattaché à cette famille, représentant une tradition différente mais pas complètement indépendante (PARKER 1978, p. 190). Les témoins M, N, P et X sont datés par la notice de la BNF du treizième siècle, sans plus de précisions, tandis que K est daté du quatorzième siècle. La datation au XIII^e s. doit être orientée vers la fin du siècle : ces ms. enregistrent des chansons de trouvères qui ne sont actifs que dans la seconde moitié du siècle, comme P, qui contient des chansons d'Adam de la Halle. Pour tous ces manuscrits, y compris K, PARKER (1978, p. 183) indique la seconde moitié du XIII^e s. Pour N, la notice de l'ARLIMA précise « fin du XIII^e s. ».⁴⁸ Leurs noms (lorsqu'ils ont un nom traditionnel) et cotes à la BFN sont les suivants :

- M BNF fr 844 ;
- N BNF fr 845 ;
- P BNF fr 847 ;
- X *Chansonnier Clairambault*, BNF fr 1050 ;
- K *Chansonnier de Navarre*, BNF fr 5198 (numéroté en pages et non en folios).

Le ms. V. Le ms. siglé V par la tradition correspond au ms. BNF fr 24406. Le ms. est divisé en deux parties ; selon les notices de l'ARLIMA⁴⁹ et de la BNF, la première partie date de la fin du XIII^e s. et la seconde partie date du début du XIV^e s. Selon PARKER (1978, p. 184), il s'agit d'un manuscrit « peu sérieux » du point de vue musical. L'auteur ajoute que V est à rapprocher de K, mais s'en distingue par une mouvance très prononcée : de nombreuses mélodies commencent comme celles de K mais évoluent ensuite sans aucun rapport, et beaucoup de pièces de V sont des *unica*, présentes dans aucune autre source (et donc pas dans K non plus).

Le ms. O. Le ms. siglé O par la tradition correspond au ms. BNF fr 846, également appelé *Chansonnier de Cangé*. Il est daté par la notice de la BNF du dernier quart du XIV^e s. O est apparenté à M et K pour les leçons de quelques chansons, mais cette parenté demeure marginale ; il présente également quelques rapprochements avec U (PARKER 1978, p. 184-185).

Le ms. a. Le ms. siglé a par la tradition correspond au ms. Reg. lat. 1490 de la Bibliothèque Apostolique Vaticane (BAV). Il dispose de parentés lointaines avec M (SCHWAN 1886 ; PARKER 1978, p. 182), mais son nœud stemmatologique terminal est constitué de manuscrits que nous n'avons pas sélectionnés. Il est daté de la première moitié du XIV^e s. par Parker et par la notice de l'ARLIMA, qui précise qu'il aurait été composé en Artois.⁵⁰

Le ms. i. Le ms. siglé i par la tradition correspond au ms. BNF fr. 12483, également appelé *Miracles de Notre Dame, en vers, etc.*, puisque les 266 premiers folios recueillent des pièces en l'honneur de la Vierge. La plupart des folios sont lacérés aux marges extérieures, les éditeurs restituent alors les notes et syllabes ou mots manquants. Le manuscrit

48: <https://www.arlima.net/no/5199>.

V : *Chansonnier français — Traité des quatre nécessaires* — Richard de Fournival, *Bestiaire d'Amour — Chansons à la Vierge* (BNF fr 24406). URL : <http://gallica.bnf.fr/ark:/12148/btv1b84386028>.

49: https://www.arlima.net/mss/france/paris/bibliotheque_nationale_de_france/francais/24406.html.

O : *Chansonnier, dit Chansonnier Cangé* (BNF fr 846). URL : <https://gallica.bnf.fr/ark:/12148/btv1b600950p>.

a : *Chansonnier français a* (BAV Reg. lat. 1490). URL : https://digi.vatlib.it/view/MSS_Reg.lat.1490.

50: <https://arlima.net/no/4833>.

i : *Miracles de Notre Dame, en vers, etc.* (BNF fr 12483). URL : <http://gallica.bnf.fr/ark:/12148/btv1b8454680s>.

date du ^{xiv}^e s. ; la notice de la BNF indique qu'il aurait été copié entre 1328 et 1373, mais la notice de l'ARLIMA⁵¹ indique que la datation peut être précisée au second quart du ^{xiv}^e s.

51: <https://www.arlima.net/no/2325>.

2.3.1.2 *Aucassin et Nicolette*

Le texte appelé *Aucassin et Nicolette*, également présent dans le NCA et dans la BFM, nous est transmis par un unique témoin : le ms. BNF fr 2168. Le texte étant un classique de la littérature médiévale, sa bibliographie est trop importante pour être ici détaillée. Toutes les références de cette section qui ne sont pas paginées proviennent de la synthèse de SARGANT-BAUR et COOK (1981). Le texte est daté du ^{xii}^e s. (PARIS 1879), de la fin du ^{xii}^e s. (FOERSTER 1904), de la fin du ^{xii}^e s. ou du début du ^{xiii}^e s. (MEYER-LÜBKE 1910), ou de la deuxième moitié du ^{xiii}^e s. (BLAKEY 1968) et est composé en dialecte picard (MEYER-LÜBKE 1910 ; DUFURNET 1984, p. 38), quoique des hypothèses wallonnes aient pu être avancées (FOERSTER 1904). Sur ce point, MEYER-LÜBKE (1910) et Dees sont en accord : Meyer-Lübke voit le texte naître à Arras, et Dees localise le texte dans le Sud-Est du Pas-de-Calais, avec une fiabilité satisfaisante (coefficient de dialectalité : 80).

Jean Bodel, *Le Vilain de Farbu*, etc. (BNF fr 2168). URL : <https://gallica.bnf.fr/ark:/12148/btv1b525135854>.

SARGANT-BAUR et al. (1981). *Aucassin et Nicolette: a critical bibliography*.

PARIS (1879). « *Aucassin und Nicolette* (Suchier) / *Aucassin et Nicolette* (Bida et Paris) [compte rendu] ».

FOERSTER (1904). « Randglossen zur Cante-fable ».

MEYER-LÜBKE (1910). « *Aucassin und Nicolette* ».

BLAKEY (1968). « *Aucassin et Nicolette*, xxiv, 4 ».

Aucassin et Nicolette est un prosimètre : y alternent régulièrement de courtes sections versifiées et chantées et de plus longues sections en prose. Chaque section est introduite par l'une des deux annotations suivantes, de la plume du scribe : *Or se cante* (« maintenant, on chante ») introduit les sections impaires, tandis que *Or dient et content et fablent* (litt. « maintenant ils disent, racontent et narrent »); ces annotations sont visibles en fig. 2.18, au-dessus de la première portée et à la fin de la dernière portée. Ainsi, *Aucassin et Nicolette* n'est pas qu'un simple prosimètre, c'est un prosimètre chanté, premier du genre dans le domaine oïlique, ce qui en fait une œuvre sans précédent, dans laquelle le chant et le récit en prose se mêlent et se complètent dans une « hybridation fonctionnelle » (PENSOM 1989, p. 64). Les parties versifiées sont des laisses assonancées de sept syllabes métriques, dont quatre sur vingt-et-une sont composées de vers féminins. Chaque laisse se termine par un vers orphelin de quatre syllabes métriques, qui échappe à l'assonance de la laisse. La musique des laisses est composée de deux phrases musicales correspondant au corps de la laisse, auxquelles s'ajoute une courte phrase correspondant au vers orphelin. Toutes les laisses ont la même musique, et la plupart des laisses sont introduites, dans le manuscrit, par la notation de la musique des deux premières phrases, en notes carrées sans rythme, sur des portées de 4 à 5 lignes ; le texte correspondant à chaque phrase musicale est noté directement en-dessous de la portée, comme visible en fig. 2.18. L'opinion la plus commune est que ces deux premières phrases musicales se répètent en alternance régulière, sous la forme d'un schéma ABAB etc. Néanmoins, quelques laisses ont un nombre de vers de sept syllabes métriques impair, ce qui s'oppose à la répétition d'un schéma ABAB continu ; mais le débat, intense, sur le schéma précis de répétition de ces phrases musicales ne peut être tranché.⁵²

PENSOM (1989). « Form and Meaning in *Aucassin et Nicolette* ».

52: On peut néanmoins consulter la section musicologique de Th. Gérold dans l'édition de ROQUES (1929b) ou VITZ (2007) pour des réflexions à ce sujet.

(ADAM DE LA HALLE 2002 [1998])⁵³ présente l'intérêt d'être diplomatique pour le texte et pour la musique. Son édition s'appuie sur le manuscrit dit de la Vallière (BNF fr 25566). Le texte, entièrement versifié, comprend des parties chantées, dont la musique est présente dans le ms., comme pour *Aucassin et Nicolette*; mais, cette fois, les notes carrées sont mensuralistes, c'est-à-dire qu'elles disposent d'une variété de formes associées à un rapport arithmétique dans les durées des notes : les *longues* ■ équivalent ici à trois *brèves* ▣ (BETTENS in ADAM DE LA HALLE 2002 [1998], p. 6), selon le principe du mode mineur parfait (cf. *infra*). Le schéma de répétition des phrases musicales est ici plus explicite, puisque la notation de la musique ne se limite pas aux premiers vers des sections musicales, comme en attestent les extraits du manuscrit et de l'édition de Bettens reproduits en fig. 2.19 et 2.20. Ici, et contrairement à *Aucassin et Nicolette*, l'alignement entre syllabes et notes est soigné, jusqu'à couper des mots entre deux lignes sur le ms.

2.3.1.4 La question rythmique

La question du rythme de la musique des trouvères est « épineuse » et « controversé[e] » (ROSENBERG, TISCHLER et GROSSEL 1995, p. 19). Le débat peut être résumé ainsi : face à des notations qui ne transmettent pas de rythme mesuré, c'est-à-dire reposant sur un rapport arithmétique de division des valeurs des notes, faut-il supposer que ces chansons ne disposent pas d'un rythme mensuraliste ? Ou bien qu'elles disposent d'un rythme mensuraliste non noté et que celui-ci doit être restitué par l'éditeur ? Nos deux ouvrages de référence pour le répertoire des trouvères témoignent de ces deux positions : ROSENBERG, SWITTEN et LE VOT (1998) s'en tiennent, la plupart du temps, à éditer les chansons en 'notes rondes' sans indication de durée relative, tandis que ROSENBERG, TISCHLER et GROSSEL (1995) éditent les chansons selon des modes rythmiques semblables, dans leurs principes, aux mètres antiques (succession de cellules composées de *longues* et/ou de *brèves*).

Le débat, dans le monde de la philologie musicale médiévale, tient en partie à l'évolution de la notation musicale au XIII^e s. : alors que les notations antérieures indiquent des rythmes qui ne correspondent pas à des subdivisions arithmétiques, les notations se complexifient au XIII^e s., notamment en lien avec le développement de l'écriture de la polyphonie (ROSENBERG, TISCHLER et GROSSEL 1995, p. 20), celle-ci nécessitant un système de transcription mensuraliste, dans lequel différents signes indiquent des durées qui correspondent à la moitié ou à un tiers de la durée indiquée par un autre signe, à la manière de la notation moderne. Ainsi, la *virga* ◀ du ms. U, signe par défaut sans valeur mensuraliste propre, encore proche en cela de la *virga* paléofranque originelle †, correspond à la note carrée dotée d'une hampe ■ dans la notation carrée. Celle-ci peut soit avoir une valeur par défaut non mensuraliste, soit, dans un système qui oppose les formes pour noter des rapports arithmétiques de durée, avoir la valeur d'une *longue*. Dans ce cas, la *longue* est décomposable en deux ou trois *brèves*,⁵⁴ notées par un *punctum* ■ (■ = ■■■ ou ■■). Les *brèves* sont elles-mêmes décomposables en deux ou trois *semi-brèves*,⁵⁵ notées par des *puncta inclinata* ◆, (■ = ◆◆◆ ou ◆◆), lesquelles deviendront avec l'*ars nova* du XIV^e s. divisibles en deux ou trois *minimes*,⁵⁶ notées † (◆ = †† † ou ††).

ADAM DE LA HALLE (2002 [1998]). *Li Gieus de Robin et de Marion*.

53: Nous citons la version en PDF de 2002 plutôt que la version en ligne de 1998, pour sa pagination.

Manuscrit de la Vallière (BNF fr 25566). URL : <https://gallica.bnf.fr/ark:/12148/btv1b6001348v>.

54: Le rapport arithmétique entre *longues* et *brèves* est appelé 'mode mineur' ; le mode mineur est parfait lorsqu'une *longue* correspond à trois *brèves*, imparfait lorsqu'une *longue* correspond à deux *brèves*.

55: Le rapport arithmétique entre *brèves* et *semi-brèves* est appelé 'temps' ; le temps est parfait lorsqu'une *brève* correspond à trois *semi-brèves*, imparfait lorsqu'une *brève* correspond à deux *semi-brèves*.

56: Le rapport arithmétique entre *semi-brèves* et *minimes* est appelé 'prolation' ; la prolation est parfaite lorsqu'une *semi-brève* correspond à trois *minimes*, imparfaite lorsqu'une *semi-brève* correspond à deux *minimes*.

39

I plerms
 Et ie men uai
 Ne plus parole ni duai
 Car ie nai mestier q' me fier
S uios
 He diez ie ne mengai p'us tierce
 Et s'est ia plus nomme de iou
 Et si ne puis auoir serou
 Se ie ne toi ou toi ou malque
 Je men uois ia faire me talque
 Ne ie nai chi plus tiens que faire
R ogans
 Barner
D aruers
 Que
R ogans
 V eus tu bien faide
 A lons uers auete a le foud
D aruers
 S or mais anchors ucel aler toue
 Au dehas ait qui ni uendra
C hi comence li gies de robin et
 de marion cadans fist. **M**arions.



obins ma
 me robins ma
 robins ma demaudee si mara. Ro

bins macara cotele de scarlatte borne
 et tele soulaame et chantu de aleu
 uia robins maime robins ma ro
Li cheua
 liers.
 bins ma demaudee si mara
 Je me repauidie du tournoiemet
 li trouuai marote seulette au cors
Ma
 gent ons **L**e robin se ta maines
 par amouus mainement **L**i chlis.
 B egiere diez uous comit ton iou
Marions
 D iez uous gart sure
Li cheualiers
 P ain touche puche de me zies
 p our toi ceste canchon amtes
 S iuolentiers et si souuent
 H e robin se tu maines
 p ar amouus mainement
Marions
 B iaus sure il ia bien pour toi

Source gallica.bnf.fr / Bibliothèque nationale de France. Département des Manuscrits. Français 25566

FIGURE 2.19 – Incipit du *Gies de Robin et de Marion*, ms. BFN fr 25566, fol. 39r. Url: <https://gallica.bnf.fr/ark:/12148/btv1b6001348v/f85>.

Marions ·



1 Robins m'ai/me robins m'a/



Robins m'a/ demande-e si m'ara :



Ro- [b] -bins m'acata cote-le ·



d'escarla- te bonne/ et bele



5 souskanie et chainture-le



a-leur/-i-va



Robins m'aime robins m'a



ro/bins m'a demande-e si m'ara /

Li cheva/liers ·



Je me repairoi-e du tournoi-ement ·/



10 Si trouvai marote seulet · au cors/ gent

Ma/ri/ons



Hé robin se tu m'aimes/



Par amours maine m'ent

Li ch[eva]liers ·

Bergiere diex vous doinst bon jour

FIGURE 2.20 – Incipit du *Gieus de Robin et de Marion*, éd. Bettens (ADAM DE LA HALLE 2002 [1998], p. 14-15).

Le système graphique évolue donc vers un système capable d'exprimer un rythme mensuraliste, lequel se réduit le plus souvent, dans le répertoire vernaculaire monodique, à l'expression d'un mode rythmique — succession plus ou moins régulière de rythmes longs et brefs, analogue aux mètres de la poésie antique occidentale. Mais les chansons des trouvères témoignent fort peu de l'application concrète de ces modes rythmiques. D'une part, le plus ancien manuscrit, U, n'est pas en notation carrée et sa notation messinne ne permet tout simplement pas l'expression d'un mode rythmique ou d'une information mensuraliste. Le seul signe musical de base du ms. U est la *virga* ◊ et ce signe ne s'oppose pas, en tant que tel, à d'autres signes isolés, comme illustré en fig. 2.21. D'autre part, les autres manuscrits, en notation carrée, ne semblent pas faire usage des potentialités mensuralistes de cette notation : pour la plupart, ils n'utilisent eux aussi qu'un seul signe de base, ■, qui correspond à la *longue* de la notation mensuraliste, comme illustré en fig. 2.22. Néanmoins, certaines chansons, dans certains manuscrits, témoignent bien d'un système mensuraliste : ainsi, dans la chanson « *Apris ai qu'en chantant plour* » du ms. N, on observe bien une opposition entre *longues* ■ et *brèves* ◻, et cette opposition exprime un mode iambique, constitué par une succession d'unités *brève - longue* (◻ ■), comme illustré en fig. 2.23 ; mais cet usage mensuraliste est marginal. Sur ces figures, le premier signe de chaque ligne est une clef d'Ut sous la forme d'un C capital plus ou moins stylisé. L'écrasante majorité des chansons est donc notée dans un système qui pourrait noter un rythme mensuraliste, mais qui ne le fait pas.

Face à cette situation, les pratiques analytiques et éditoriales des philologues et musicologues se divisent donc en deux camps : il y a ceux qui se refusent à projeter une information rythmique absente des sources, et ceux qui se refusent à ne pas le faire. Cette seconde posture est particulièrement problématique à des fins d'analyse, dans la mesure où rien ne vient consolider définitivement une analyse rythmique. Ainsi, DRAGONETTI (1960, p. 524 et sq.) expose que deux musicologues ont pu parvenir à des interprétations rythmiques différentes des mêmes chansons. Une telle difficulté ne semble pas gêner outre-mesure les tenants de l'analyse en modes rythmiques, comme en témoignent Rosenberg, Tischler et Grossel :

La possibilité d'arriver à des résultats différents n'invalide cependant pas l'application des modes rythmiques aux chansons des trouvères ; en fait, cela souligne la facilité avec laquelle le musicien médiéval pouvait varier, selon les circonstances, l'exécution d'une pièce.

(ROSENBERG, TISCHLER et GROSSEL 1995, p. 22)

Cette citation appelle deux commentaires, qui fondent notre rejet de la donnée mensuraliste ('modes rythmiques') dans l'analyse de la musication des chansons de ce répertoire. Le premier commentaire porte sur la méthode, la première partie de la citation mettant en lumière un problème épistémologique fondamental : s'il n'est pas possible de déterminer avec certitude le mode rythmique d'une pièce, alors celui-ci ne peut faire partie des données de l'analyse. Le second commentaire porte sur la deuxième partie de la citation, et sur la possibilité, pour un musicien, d'opter plus ou moins librement pour un mode rythmique plutôt que pour un autre. Si tel est bien le cas, deux hypothèses sont

DRAGONETTI (1960). *La technique poétique des trouvères dans la chanson courtoise : contribution à l'étude de la rhétorique médiévale.*

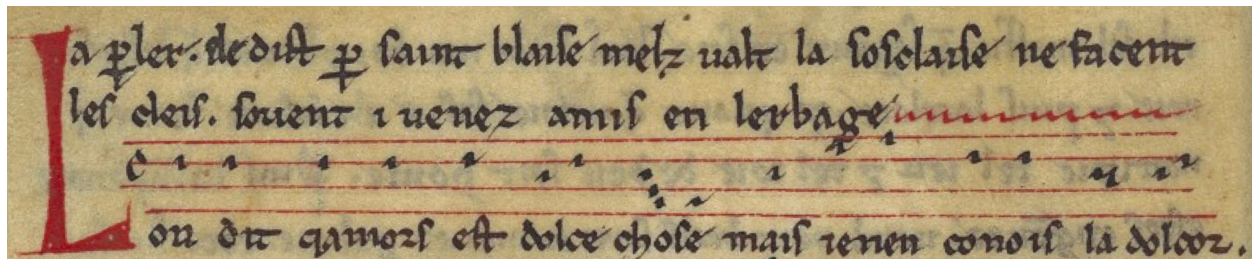


FIGURE 2.21 – Exemple de notation impropre au rendu d'un mode rythmique : « *L'on dit qu'Amors est dolce chose* », pièce notée dans le ms. U (fol. 47v.). Url : <https://gallica.bnf.fr/ark:/12148/btv1b60009580/f102>.

La notation messine de l'ensemble du manuscrit ne permet pas de noter les durées des notes, puisqu'un seul signe de base existe : la *virga* ◄. Les groupes de signes, tels la *clivis* ♪ ou le *climacus* ♫, ne transmettent pas d'information rythmique de type mensuraliste.



FIGURE 2.22 – Exemple de notation qui pourrait rendre un rythme mensuraliste ou modal mais ne le fait pas : « *Au renouvel du tens que la florete* », pièce notée dans le ms. P (fol. 188r.). Url : <https://gallica.bnf.fr/ark:/12148/btv1b8454673n/f387>.

Les notes carrées avec une hampe vers le bas correspondent aux *uirgae* du ms. U en fig. 2.21 : elles représentent une durée non contrastive.



FIGURE 2.23 – Exemple de notation rendant un mode rythmique : « *Apris ai qu'en chantant plour* », pièce notée dans le ms. O (fol. 10 v.). Url : <https://gallica.bnf.fr/ark:/12148/btv1b6000950p/f48>.

Ici, le mode est iambique : les notes simples pourvues d'une hampe ◼ sont des *longues*, qui valent deux *brèves* notées sans hampe ◻, qui valent deux *semi-brèves* notées par les *puncta inclinata* ◄ dans les signes complexes ; le signe composé de la seconde ligne, la *clivis* ♪, vaut une *longue*.

possibles : soit ces modes rythmiques disposent de dénominateurs communs qui sont, seuls, pertinents pour la bonne-formation de l'association entre notes et syllabes, auquel cas l'objet pertinent de l'analyse de la musication ne serait pas le mode rythmique lui-même mais un niveau d'abstraction supplémentaire, regroupant les modes par familles. Soit, seconde hypothèse, le mode rythmique est tout simplement non pertinent pour la bonne-formation de l'association entre notes et syllabes dans ce répertoire. Cette possibilité de *changer* le mode rythmique d'une pièce s'inscrit dans un contexte plus large où texte et musique ne sont pas fixement associés l'un à l'autre, ce dont témoignent les *contrefacta* nombreux : trouvères et jongleurs (exécutants n'étant pas des compositeurs) peuvent tout à fait chanter un texte sur la mélodie d'un autre, et, réciproquement, chanter une mélodie sous un autre texte que celui auquel elle était originellement associée. Ainsi, une même chanson peut être attestée sous différents modes rythmiques dans des manuscrits différents (DRAGONETTI 1960, p. 526). Lorsque l'information mensuraliste est absente de la source, l'éditeur se repose alors sur la régularité des accents dans le texte (ROSENBERG, TISCHLER et GROSSEL 1995, p. 22-23), ce qui le force à accepter une certaine quantité de syllabes atones sur des temps forts et de syllabes toniques sur des temps faibles, les accents du texte n'étant jamais réguliers qu'en fin d'hémistiche et de vers. Il s'agit donc de repérer les positions métriques du texte les plus accentuées et d'en déduire un mode rythmique, ce qui repose sur l'idée que l'on trouve, dans le texte, une métrique organisée en pieds. Or cette régularité supposée des accents au sein du texte fait cruellement défaut : d'une part, une analyse en pieds de la poésie des trouvères fait appel à un répertoire de pieds bien trop grand (pieds composés de une à quatre syllabes chez Dragonetti) et à une liberté de succession des pieds bien trop permissive (les possibilités de succession des pieds semblent virtuellement infinies) pour que ces pieds aient une quelconque valeur proprement métrique. D'autre part, la fréquence apparente des accents au sein des vers n'est apparente qu'à la lecture de quelques vers ; lorsque l'on prend le texte dans son ensemble, les régularités statistiques s'effondrent. Ainsi, dans notre mémoire de master, nous avons annoté systématiquement les syllabes accentuables de tous les vers de 24 chansons de trouvères, et parvenons à la conclusion — certes sur un corpus restreint — que seules les positions qui terminent les constituants que sont l'hémistiche et le vers sont dotées d'une réelle régularité métrique :

Toutes les positions à l'intérieur de ces constituants peuvent être indifféremment occupées par des éléments toniques ou atones, à l'exception de la position qui précède immédiatement la tonique, qui est nécessairement atone du fait de l'algorithme de résolution des conflits accentuels.

(PREMAT 2017, p. 174)

Ainsi, toute régularité perçue à la lecture des premiers vers d'une chanson s'effondre si l'on prend la peine de sérieusement étudier l'ensemble de la chanson. Enfin, il faut noter que nous parlons ici de *syllabes accentuables* et non de *syllabes accentuées* ou d'*accents* comme le fait, p.ex., Dragonetti. De fait, le répertoire des trouvères se trouve à la bascule entre l'accent de mot et l'accent de groupe, et, précisément, au moment-même où disparaissent les régularités accentuelles internes aux constituants dans la poésie médiévale française, avec la disparition des rythmes iambiques et du marquage de la quatrième syllabe des octosyllabes (RAINSFORD 2010,

2011a, 2011b, p. 125-163). Ainsi, dans le corpus des trouvères, l'absence de régularité des syllabes dites accentuables, c'est-à-dire de la dernière syllabe non-schwa d'un mot lexical et de certains mots fonctionnels, hors des positions de fin d'hémistiche et de vers, est pleinement cohérente avec le changement métrique attesté dans des textes non chantés, et avec le changement linguistique.

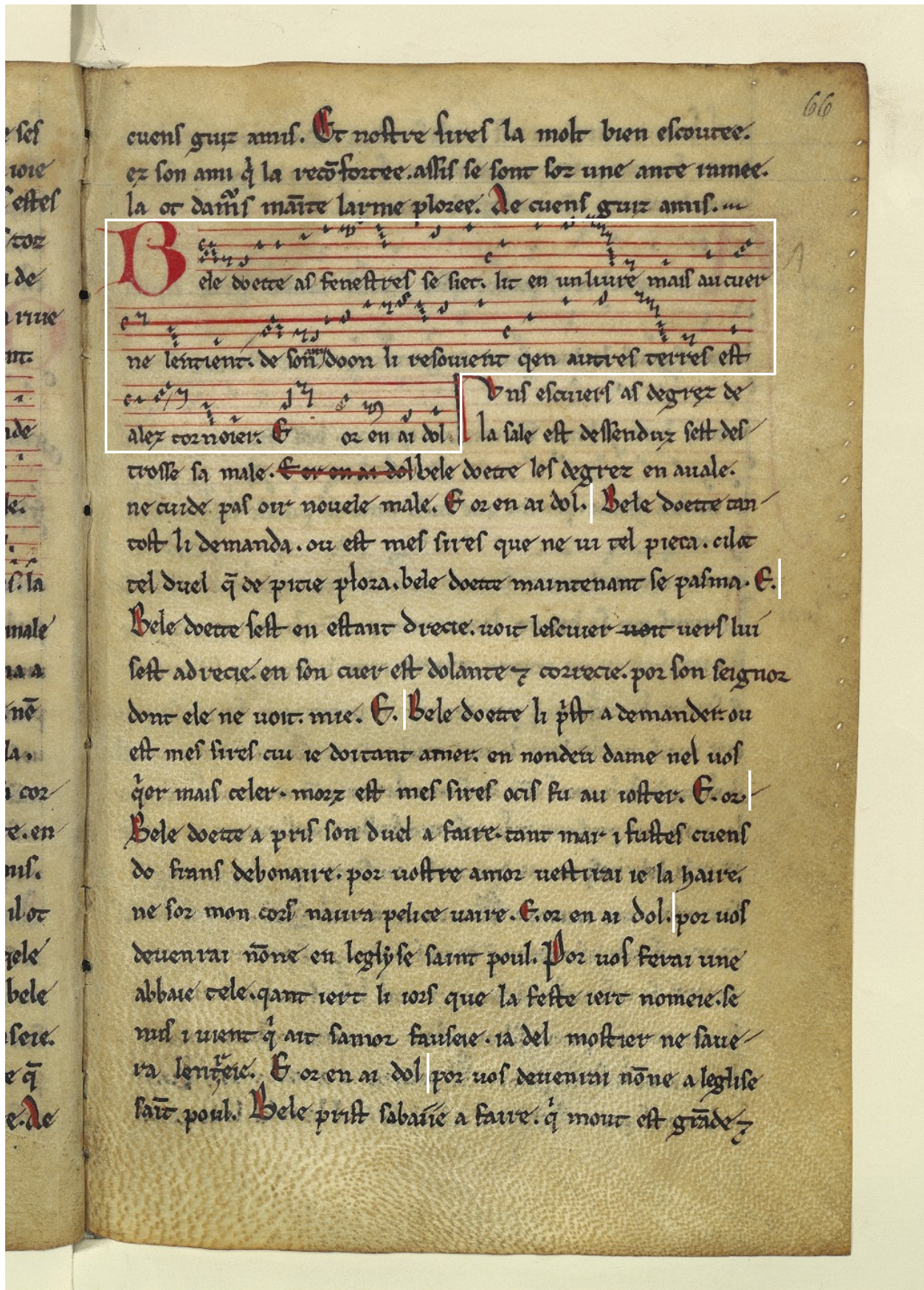
Il nous paraît difficile de tirer de tous ces arguments une pertinence structurelle des modes rythmiques, et nous adhérons donc naturellement à l'interprétation de Le Vot (ROSENBERG, SWITTEN et LE VOT 1998, p. 8) selon laquelle les chansons des trouvères doivent être, pour la plupart, considérées comme dépourvues d'un rythme mensuraliste, correspondant par là à une forme de *musica immensurata, plana sine mensurabilis* ou *sine mensura*, comme l'est le plain chant. Ainsi, dans leur ouvrage, ces auteurs n'éditent en rythmes mesurés que quelques chansons, c'est-à-dire celles dont les témoins sont explicitement mesurés (comme « *Apris ai qu'en chantant plour* » en fig. 2.23) ou, plus exceptionnellement, celles qui comportent de très forts indices rythmiques dans le texte. Notre approche générale consiste donc à ne pas analyser le rythme dans ces répertoires. Cette approche est cohérente avec notre rapport aux sources musicales : nous ne prenons tout simplement pas en considération les éléments qui ne sont pas présents sur les sources manuscrites.

Par extension, nous appliquons aussi cette approche à *Aucassin et Nicolette*, dont la notation musicale n'oppose pas de signes simples. En effet, comme la plupart des chansons de trouvère, l'unique ms. d'*Aucassin et Nicolette* fait usage d'un seul signe simple, la *virga* de la notation carrée : ▣. Nous rejetons donc l'analyse trochaïque (♩ ♪) que propose Gérold dans l'édition de ROQUES (1929b).

2.3.1.5 Musique notée et itération musicale

Dans les sources musicales médiévales, lorsqu'une musique est répétée de strophe en strophe (itération inter-strophique), la musique n'est le plus souvent écrite qu'une fois, et le seul texte qu'on trouve directement sous les portées musicales est celui de la première strophe. Cette situation, analogue à la présentation des partitions modernes, est illustrée en fig. 2.24. Ce fait induit une nouvelle limite à toute étude rigoureuse des procédés d'association entre notes et syllabes : la musique qui est notée n'est pas nécessairement la musique de tout le texte, mais peut n'être la musique que du texte de la première strophe. Ainsi, la présence d'une note pour une terminaison féminine (coupe épique ou rime féminine) ne dit rien de ce qu'il advient si, dans une autre strophe, la même musique doit être associée à une terminaison masculine, et, réciproquement, l'absence d'une note correspondant à une terminaison masculine ne dit rien de ce qu'il advient si, dans une autre strophe, la même musique doit être associée à une terminaison féminine.

En effet, la mélodie et le rythme (ici non mensuraliste) ne sont pas nécessairement identiques de phrase en phrase, la musique admettant des adaptations pour suivre la présence ou absence d'une syllabe extramétrique dans le texte (PREMAT et CHOUVION 2021). Dans la plupart des chansons, ces procédés ne sont donc attestés qu'au sein de la première strophe. Dans l'exemple de la chanson « *Bele Doette* » illustrée en



Source gallica.bnf.fr / Bibliothèque nationale de France. Département des Manuscrits. Français 20050

FIGURE 2.24 – Non répétition graphique de la musique sur la chanson « *Bele Doette* », ms. U, fol. 66 r. Url : <https://gallica.bnf.fr/ark:/12148/btv1b60009580/f139>.

Le texte et la musique de la première strophe sont encadrés en blanc ; ensuite, le changement entre chaque strophe est indiqué par une barre verticale après le vers *Et or en ai dol* qui, comme un refrain, vient terminer chaque strophe.

fig. 2.24, la répétition de la musique de strophe en strophe (itération inter-strophique) se double d'une répétition de la musique au sein de chaque strophe (itération intra-strophique) : la musique de la strophe est composée sur un schéma ABABC, où la phrase A est associée aux vers 1 et 3, la phrase B aux vers 2 et 4 et la phrase C au vers 5, qui, à la manière d'un refrain, vient clore chaque strophe. Cette fois-ci, l'itération est bien écrite sur le manuscrit : les phrases A et B sont notées deux fois chacune. On peut ainsi observer comment la musique s'adapte à la présence d'une coupe épique au premier vers (*Bele Doet<te> as fenestres se siet*)⁵⁷ et à l'absence de celle-ci au troisième vers (*A son amy Doon li resovient*; le mot *amy* figure en ajout surscrit sur le ms.).

57: Les symboles < et > indiquent que la syllabe qu'ils enserrent est extramétrique.

FIGURE 2.25 – Répétition intra-strophique et variation de la musique sur « *Bele Doette* », ms. U, fol. 66 r, v. 1 et 3.

Figure issue de PREMAT et CHOUVION (2021, p. 259); URL du ms. : <https://gallica.bnf.fr/ark:/12148/btv1b60009580/f139>; transcription moderne des auteurs, essentiellement semblable à celle de ROSENBERG, SWITTEN et LE VOT (1998).

Comme illustré en fig. 2.25, la séquence musicale à cheval entre les hémistiches est *Do Do Ré*. Le premier *Do* est toujours associé à la tonique du premier hémistiche, c'est-à-dire *-et-* de *Doette* au v. 1 et *-y* de *amy* au v. 3. Au vers 1, le second *Do* correspond à la syllabe extramétrique (<*-te*> de *Doette*), mais au v. 3 il correspond à la première syllabe du second hémistiche (*Do-* de *Doon*, *Doon* étant toujours bisyllabique) : il est rattaché à la même syllabe que *Ré*. Le changement dans la syllabité est marqué par la constitution de *Do Ré* en neume complexe (ici, un *pes rotundus* ♪, constitué de deux *virgae* en un trait de plume), qui indique que les deux notes portent sur la même syllabe ; la constituance neumatique est rendue dans nos transcriptions par des ligatures. Ce phénomène peut s'analyser comme une adaptation de la musique à la présence/absence d'une coupe épique par *spécification de la syllabité* et adaptation de la constituance neumatique (PREMAT et CHOUVION 2021, p. 259-263).

Les phénomènes de cette nature sont relativement nombreux, et témoignent bien du fait que la musique s'adapte à certaines variations de la longueur des vers. Par là, ils témoignent aussi du fait que la musique qui est notée au-dessus de chaque vers n'est que la musique de ce vers, et qu'on ne sait rien des adaptations qu'il faut prévoir pour les autres vers. Aussi, dans nos publications sur ce sujet, nous nous en sommes

tenus à n'étudier l'association entre musique et texte que pour les vers qui sont notés directement sous une ligne musicale, ce qui correspond la plupart du temps à la première strophe des chansons. Cette approche est reproduite dans ce travail.

La seule exception que nous nous autorisons à cette restriction concerne les chansons dont la musique serait notée plusieurs fois, que ce soit parce que la musique est présente dans plusieurs manuscrits (cf. [PREMAT et CHOUVION 2021](#), p. 254), ou parce que, au sein de la même source, la même musique revient plusieurs fois (cf. [PREMAT et CHOUVION 2021](#), p. 256-257). Le deuxième cas concerne, par exemple, *Aucassin et Nicolette*, dont la musique, qui est équivalente pour toutes les laisses, est notée à nouveau pour chaque laisse.

2.3.1.6 En somme

Notre étude des chansons des trouvères, d'*Aucassin et Nicolette* et du *Gieus de Robin et de Marion* se fonde toujours sur l'étude des sources manuscrites. Elle comporte un certain nombre de restrictions, qui nous semblent essentielles pour garantir la qualité de l'analyse, et qui se résument à un strict respect des données attestées dans les sources : nous ne considérons le rythme que lorsque celui-ci est effectivement noté dans la source, et ne prenons en compte que le texte des vers qui sont écrits directement sous une phrase musicale.

2.3.2 En somme

En complément aux corpus informatisés, nous recourons pour les sources musicales à des corpus manuscrits. Ces corpus sont parmi les plus anciens, avec des chansons de trouvères, des laisses chantées d'*Aucassin et Nicolette* et des parties musicales du *Gieus de Robin et de Marion*. Dans tous les cas, les manuscrits forment l'assise empirique de notre travail ; soit nous avons travaillé sans partitions modernes, soit nous avons travaillé avec des partitions modernes mais avons systématiquement contrôlé sur les témoins les informations analysées. Nous nous en tenons strictement aux informations contenues dans ces manuscrits, en rejetant toutes les informations qu'une analyse musicologique interventionniste pourrait projeter sur ces sources : ainsi, le rythme mensuraliste est rejeté pour les trouvères et *Aucassin et Nicolette*, et les corrections, qu'elles soient musicales ou textuelles, sont rejetées dès lors que le manuscrit est clairement lisible. Dans la même optique, nous ne projetons pas l'ensemble du texte sur la musique, et nous tenons à analyser l'association entre texte et musique des vers qui sont copiés directement sous une portée musicale.

2.4 Conclusion

Cette thèse repose sur un corpus principal, le NCA, issu des matériaux de DEES (1987). Ce corpus comprend 296 textes intégraux ou sous forme d'extraits, annotés en lemmes et en partie du discours, ces dernières étiquettes intégrant la flexion. Les textes disposent de métadonnées diatopiques et chronologiques multiples ; nous privilégions les localisations de Dees et les datations de composition. Le NCA ne comprend pas de textes antérieurs à 1100, et a été nettoyé des accents qui différencient un <e> final représentant une voyelle tonique d'un <e> final représentant un schwa. Aussi, pour documenter la situation du très ancien français et le traitement métrique des voyelles finales atones, nous recourons à la BFM. L'annotation y est un peu moins riche (absence d'information sur le genre et le nombre), mais une partie de cette annotation a été revue à la main et est très qualitative. Enfin, nous complétons cette analyse métrique par une analyse du traitement des voyelles finales atones dans la musique vocale de langue d'oïl, à partir d'un corpus allant du ^{xiii}e s. à la bascule entre les ^{xiv}e et ^{xv}e s., et nous avons détaillé quelles étaient les limites philologiques inhérentes à ce corpus et comment nous pouvions les prendre en compte dans l'analyse.

La partie de linguistique informatique de cette thèse repose principalement sur trois outils : TXM, Comparalem et le PAM. Le premier est utilisé pour accéder aux corpus et pour en extraire les données qui formeront l'*input* de Comparalem. Le deuxième sert au repérage des lemmes ayant une variation graphique <e>/Ø (où <e> représente un schwa), tandis que le troisième sert au relevé des régularités et irrégularités métriques imputables au traitement du schwa final. Dans ce chapitre, nous traitons successivement de ces trois outils.

3.1 TXM

TXM (HEIDEN, MAGUÉ et PINCEMIN 2010) est une plateforme de construction, d'interrogation et de consultation de corpus linguistiques. Cette plateforme est développée par Matthieu Decorde, Serge Heiden, Sébastien Jacquot, Alexei Lavrentiev et Bénédicte Pincemin. Elle est disponible en installation locale (sur Linux, Mac et Windows) ou par le biais d'une interface Web. Nous utilisons principalement la version 0.8.1 en installation locale. La version 0.8.2 est disponible depuis le 10 juin 2022 ; nous ne l'avons pas adoptée pour conserver un flux de travail constant. La principale fonction de TXM que nous utilisons est celle des requêtes de type Index, qui nous permettent d'extraire les données qui serviront d'*input* à Comparalem.

HEIDEN et al. (2010). « TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement ».

3.1.1 Requêtes CQP de type Index

TXM permet des analyses textométriques à partir de corpus textuels, annotés ou non, selon une architecture open-source et modulaire. Dans notre cas, les deux corpus sur lesquels nous travaillons, la BFM et le NCA, ont été importés dans TXM directement sous la forme de fichiers binaires ; nous n'avons donc pas construit nous-même les corpus avec TXM. Ces corpus ont été importés dans TXM avec leur annotation déjà existante ; nous n'avons donc pas utilisé TXM pour annoter les corpus. Sur ces corpus annotés, TXM permet l'exécution de requêtes (en syntaxe CQP) portant sur les chaînes de caractères et/ou sur les propriétés associées à ces chaînes. Il est par exemple possible de demander à TXM de produire la liste de tous les mots se terminant par un <e> graphique et ayant pour propriété morphologique d'être un nom commun. Les requêtes de TXM permettent également de prendre en compte plusieurs mots, ce qui permet, dans notre cas, d'interroger la variation <e>/Ø en fonction de la présence ou absence d'une consonne en tête du mot suivant. Cette fonction permet notamment l'export des fichiers qui serviront d'*input* au programme Comparalem. Pour cet usage, les requêtes ont été raffinées par étape, de manière à améliorer la spécificité du test utilisé en éliminant les faux positifs, parfois au prix d'une perte modérée de sensibilité.

Ce type de requêtes et leur raffinement progressif est illustré en (3.1-4). La requête (3.1) renvoie l'ensemble des noms se terminant par un graphème <e>. Les caractères « . » et « * », utilisés en syntaxe CQP comme dans les expressions régulières, servent ici à indiquer la position des chaînes recherchées : « . » renvoie à n'importe quel caractère alphabétique, et « * » indique que ce caractère peut être présent de zéro à une infinité de fois ; ainsi, placer « . * » avant la chaîne recherchée permet de signaler que la chaîne recherchée est en fin de mot (word = ". *e"), et placer « . * » après la chaîne recherchée permet de signaler que la chaîne recherchée est placée en début de mot (pos = "NOM. *") inclut toutes les indications de flexion de l'étiquette pos, p.ex. NOM_obj_femi_sg pour un nom féminin singulier au cas régime). Les différentes propriétés d'une même forme sont additionnées (de manière cumulative) par l'opérateur &, et l'ensemble des instructions portant sur une forme est délimité par les crochets [et].

- (3.1) [word=". *e" & pos="NOM. *"]
 Glose : cherche les mots qui se terminent (. * [...]) par un <e> et qui ont une étiquette pos qui commence ([...] . *) par la chaîne « NOM ».
 ⇒ 173 198 occurrences dans le NCA.

Cette première requête comporte des mots dont le <e> final ne représente pas un schwa mais une voyelle /e/, comme cite (« cité »), degre (« degré »), divinite (« divinité »), etc. Dans une écrasante majorité des cas, le lemme de ces mots se termine par un <é> ; on peut donc exclure une grande partie de ces <e> finaux pour /e/ en excluant les lemmes qui se terminent par <é>. L'exclusion est réalisée par l'ajout d'un opérateur « ! » devant le signe « = » (3.2). Cette restriction permet d'exclure 7191 occurrences de noms communs se terminant par <e> pour /e/.

- (3.2) [word=". *e" & pos="NOM. *" & lemma!=" . *é"]
 Glose : idem (3.1) mais dont le lemme ne se termine pas par le caractère « é ».
 ⇒ 166 007 occurrences dans le NCA.

Si l'on veut recueillir uniquement les occurrences de noms communs se terminant par <e> pour /ə/ situés devant un mot à initiale vocalique, on peut ajouter dans la requête un second mot, délimité par un nouveau jeu de crochets. Dans la mesure où les requêtes CQP n'ont pas de classes opposant les voyelles et les consonnes, il faut ici faire une liste des différents graphèmes pouvant correspondre à des voyelles ; cette liste se fait sous la forme d'un ensemble d'alternatives ; l'ensemble est placé entre parenthèses, et les alternatives sont séparées par un caractère « | ». En (3.3), nous spécifions que le nouveau mot commence par une voyelle « a », « e », « i », « o » ou « u ». Dans la mesure où les initiales vocaliques sont bien plus rares que les initiales consonantiques en ancien français, le nombre d'occurrences est divisé par quatre.

- (3.3) [word=". *e" & pos="NOM. *" & lemma!=" . *é"]
 [word="(a|e|i|o|u) . *"]
 Glose : idem (3.2) et suivis d'un mot commençant par une voyelle (« a » ou « e » ou « i » ou « o » ou « u »).
 ⇒ 43 039 occurrences dans le NCA.

La spécification du contexte initial du second mot en (3.3) est imparfaite : elle contient deux graphèmes, <i> et <u>, qui peuvent représenter une voyelle (p.ex. *il, unques*) ou une consonne (p.ex. *ieuner* : « jeûner » ; *iaw* : « eux » ; *ual* : « val, vallée »). En (3.4), nous avons exclu les graphèmes <i> et <u> des graphèmes initiaux du second mot ; nous avons choisi d'appliquer systématiquement cette exclusion, pour ne pas introduire de faux positifs chaque fois qu'une consonne initiale est notée par <i> ou <u>. Cette exclusion a pour effet une perte de sensibilité du test, mais il nous a semblé préférable d'augmenter la spécificité du test aux dépens de sa sensibilité. Nous avons fait de même pour le graphème consonantique <v> dans la spécification des initiales consonantiques ; dans la mesure où Dees n'a pas unifié les graphies <u/v>, <v> peut parfois noter une voyelle (p.ex. *vne* : « une »). Ici, le résultat de l'exclusion de <i> et <u> initial est faible (529 occurrences sont retirées), mais sur d'autres parties du discours il peut représenter jusqu'à un quart des occurrences, notamment lorsque la partie du discours interrogée est souvent suivie du pronom personnel sujet noté *ie* ou *io*.

```
(3.4) [word=".*e" & pos="NOM.*" & lemma!=".*é"]
      [word="(a|e|o).*" ]
      Glose : idem (3.3) mais à l'exception de « i, u » à l'initiale du second
      mot.
      ⇒ 42 510 occurrences dans le NCA.
```

De telles requêtes interrogent donc à la fois la forme et les propriétés attachées à celle-ci. Elles permettent d'isoler une partie du discours, voire d'isoler des formes fléchies au sein d'une partie du discours. On peut ainsi isoler les déterminants définis selon qu'il sont masculins ou féminins, singuliers ou pluriels, au cas régime ou au cas sujet. Pour l'utilisation de Comparalem (cf. section suivante), les requêtes sont la plupart du temps formulées de manière à capturer d'une part les formes qui se terminent par un <e> pour schwa sans coda, et d'autre part les formes qui se terminent par une voyelle autre que <e> pour schwa, toujours sans coda. Parfois, dans le cas d'alternances qui ne se résument pas à l'opposition entre <e> pour schwa et une voyelle finale autre que <e> pour schwa, les requêtes portent sur l'opposition entre la présence d'une voyelle finale et l'absence de celle-ci (p.ex. pour distinguer les formes élidées de l'article *l'* des formes non élidées *li, le, lo*, etc.). Toutes les formes sont vérifiées dans les textes pour contrôler la spécificité des requêtes, et lorsque des formes non pertinentes sont relevées, soit les requêtes sont réécrites soit, dans le cas de formes qu'on ne peut pas exclure des requêtes sans affecter trop la sensibilité du test, sont exclues manuellement par l'usage d'une liste d'exceptions.

3.1.2 En somme

Ainsi, TXM nous sert à accéder aux corpus (pour lire les textes, vérifier des occurrences, etc.) et à extraire les jeux de données qui serviront d'*input* au programme Comparalem. Cette extraction se fait au moyen de requêtes du type « Index », dont la syntaxe CQP permet d'interroger la forme et les propriétés d'annotation.

3.2 Comparalem

Comparalem est le nom que nous avons donné à un ensemble de scripts en langage R que nous avons développés pour cette thèse. Il peut s'exécuter dans n'importe quel environnement permettant l'exécution de R; pour cette thèse, il a été exécuté dans le logiciel R Studio. Ce programme vise à obtenir une détection semi-automatique des lemmes dotés d'une variation entre <e> pour schwa et \emptyset (absence de graphème) en position finale. Son principe général est le suivant : Comparalem prend en *input* deux ensembles de formes, associées à leurs lemmes, à leurs étiquettes morphosyntaxiques et à l'identifiant du texte dans lequel elles apparaissent. L'un de ces deux ensembles correspond aux formes dotées d'un schwa final sans coda, l'autre correspond aux formes sans schwa final. Le programme compare ces deux listes pour trouver les lemmes communs entre ces deux listes. Le fonctionnement conceptuel du programme est résumé en fig. 3.1. Comparalem applique ensuite un certain nombre de mesures statistiques et produit des sorties graphiques permettant de visualiser la distribution des formes des lemmes repérés dans les deux listes, en fonction du temps et de l'espace.

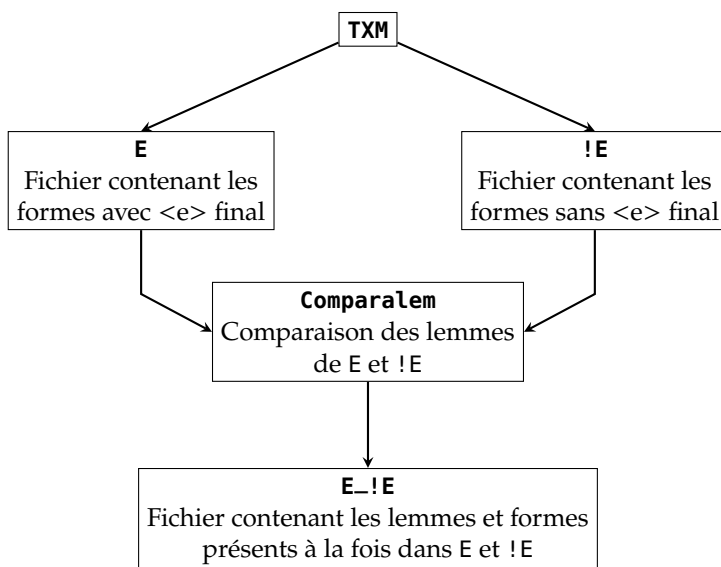


FIGURE 3.1 – Fonctionnement simplifié de Comparalem. Figure issue de PREMAT (2023).

Dans sa version développée pour cette thèse, Comparalem est orienté vers la détection de l'alternance <e>/ \emptyset finale, mais son architecture lui permet en réalité de travailler sur n'importe quelle variation graphique — nous avons également utilisé ce programme pour la détection de l'alternance <e>/ \emptyset en position interne dans le cas de la syncope et de la réduction des hiatus, en chap. 11. En effet, la variation étudiée par Comparalem ne dépend que d'une seule chose : des tables d'occurrences annotées qui lui sont données en *input*. Celles-ci, produites par TXM à partir de requêtes de type Index (cf. sec. précédente), peuvent très bien concerner n'importe quelle variation graphique, et Comparalem s'en tient à comparer les lemmes présents dans les deux fichiers. La spécialisation du programme, dans sa version actuelle, au traitement des alternances <e>/ \emptyset finale ne tient qu'au nom donné aux objets dans le code informatique, aux commentaires du script et aux étiquettes dans les sorties graphiques ;

une version publique débarrassée de cette spécialisation artificielle sera mise au point et le programme sera accessible librement dans le futur.

3.2.1 Flux de travail de Comparalem

Dans le détail, le fonctionnement du programme est plus complexe que ce qui est résumé en fig. 3.1. Le programme s'exécute en deux temps, sur des jeux de données différents : dans un premier temps, le programme est exécuté sur les formes sans leur contexte (indifféremment de l'initiale du mot suivant), de manière à créer la liste des lemmes disposant de cette variation <e>/Ø (finale, ou interne dans le cas des syncopes et de la réduction des hiatus). Cette liste est exportée dans un fichier tabulaire, appelé R [POS] merged freq, où [POS] est remplacé par la partie du discours telle que l'utilisateur veut la désigner. C'est cette première étape qui est représentée en fig. 3.1. Dans un second temps, le programme est exécuté à nouveau, cette fois-ci avec comme *input* les requêtes TXM contextuelles (selon l'initiale du mot suivant) ainsi que le fichier R [POS] merged freq. À chaque étape, l'utilisateur peut éditer le fichier R [POS] merged freq ainsi qu'un fichier d'exceptions, nommé [POS]_irrelevant forms, qui lui permettent d'exclure des lemmes et des formes, principalement à cause de défauts dans l'annotation des corpus ou d'idiosyncrasies lexicales qui n'ont pas pu être prises en compte dans la rédaction des requêtes CQP dans TXM. Avant l'exécution de Comparalem, l'utilisateur doit donc procéder à une étape de préparation des données dans TXM. Dans ce qui suit, nous abordons les éléments pertinents de chacune de ces trois étapes, dans leur ordre d'exécution.

3.2.1.1 Étape préliminaire : préparation des données

La première étape du flux de travail que nous utilisons s'effectue en amont du programme Comparalem, par la préparation de jeux de données idoines issus de TXM. Cette préparation s'effectue au moyen de requêtes CQP, telles que décrites *supra* dans la section consacrée à TXM (sec. 3.1). Les tables lexicales issues des requêtes de type Index de TXM doivent comporter un certain nombre d'informations, dans un ordre fixe, tel que représentées en (3.5).

- (3.5) Informations qui doivent être incluses dans les tables lexicales
- (a) Forme de l'occurrence (word)
 - (b) Partie du discours de l'occurrence (pos)
 - (c) Lemme (lemma)
 - (d) Identifiant du texte et de la forme au sein du texte (id)
 - (e) Fréquence (F)

L'export de ces informations depuis TXM produit un fichier CSV à deux colonnes, tous les attributs sauf la fréquence étant concaténés dans une même colonne, et séparés entre eux par un tiret bas « _ ». Par ailleurs, au sein de ces attributs concaténés, le nombre d'en-têtes ne correspond pas au nombre de variables que l'on veut obtenir : pos correspond à la partie du discours mais aussi aux informations flexionnelles, id correspond à l'identifiant du texte et à l'identifiant unique de la forme, ce dernier étant lui-même bipartite, et un w est inséré entre le lemme et l'identifiant,

généralisant une variable sans en-tête. Ces fichiers ont donc été retravaillés par des commandes SED¹ exécutées dans un terminal de manière à obtenir une colonne par variable. Ces commandes ont fait l'objet d'une rédaction *ad hoc* pour chaque jeu de données, mais leur fiabilité est avérée dans la mesure où l'import des données dans Comparalem échoue systématiquement en cas de malformation des fichiers. L'effet de ces commandes, dans le détail desquelles nous ne rentrerons pas ici, est illustré par la transformation de (3.6) en (3.7).

1: *Stream Editor* (en.) : éditeur de flux.

(3.6) Exemples de formes adjectivales en sortie de TXM

```

1 | word_pos_lemma_id                F
2 | abondeuse_ADJ_suj_femi_sg_UNKNOWN_w_calen_s50638_4    1
3 | acceptable_ADJ_obj_femi_sg_acceptable_w_fablesA_s131573_5  1
4 | acointable_ADJ_suj_femi_sg_acointable_w_clef_s72537_6    1

```

(3.7) Fichier CSV obtenu après exécution de commandes SED

```

1 | word      ; pos ; decl ; gender ; number ; lemma      ; id      ; ref1   ; ref2 ; F
2 | abondeuse ; ADJ ; suj  ; femi  ; sg    ; UNKNOWN   ; calen   ; s50638 ; 4     ; 1
3 | acceptable ; ADJ ; obj  ; femi  ; sg    ; acceptable ; fablesA  ; s131573 ; 5     ; 1
4 | acointable ; ADJ ; suj  ; femi  ; sg    ; acointable ; clef    ; s72537  ; 6     ; 1

```

Dans quelques cas, toujours mentionnés dans l'exposé des données, nous avons également utilisé des commandes SED pour éditer les lemmes dans les données. C'est par exemple le cas lorsque nous travaillons sur les pronoms personnels, dans la mesure où TreeTagger propose trois lemmes correspondant à la même unité : « je », « je-moi » et « moi ». Non seulement ces étiquettes ne sont pas exclusives, mais la notion de lemme est ici problématique. Nous avons ainsi préféré relemmatiser ces formes sous la forme d'une étiquette abstraite « PRO_1sg ». Lorsque la relemmatisation crée un nouveau lemme qui ne se démarque pas immédiatement des lemmes de Tree Tagger, nous avons toujours ajouté « _relemma » à la fin du lemme pour signaler son origine. C'est par exemple le cas pour les multiples étiquettes de lemmes du mot *coart*, Tree Tagger ayant visiblement des hésitations quant au diacritique marquant l'hiatus : il dispose, pour le même lemme au sens linguistique, de trois lemmes informatiques (« coart », « cöart » et « coärt »), et peut les combiner entre eux (« coart-coärt »). Ces lemmes et combinaisons de lemmes ont été remplacés par l'étiquette « coart_relemma ».

En somme. La préparation des données se fait, en amont de l'usage de Comparalem, dans TXM, par des requêtes CQP de type Index. Les données, exportées en CSV, sont ensuite transformées par des commandes SED pour obtenir des variables indépendantes et cohérentes, et, ponctuellement, pour intervenir sur la lemmatisation.

3.2.1.2 Première étape : générer la liste des lemmes à variation

La première étape concernant Comparalem à proprement parler, tout comme la seconde, se déroule dans un environnement de développement intégré R, par l'exécution des scripts formant le programme Comparalem. Elle est, au niveau du code informatique, assez rudimentaire : il s'agit d'obtenir la liste des lemmes présents dans les deux fichiers d'*input* et de

l'exporter sous la forme du tableau R [POS] merged_freq. De manière à ce que l'utilisateur puisse faire un usage raisonné de ce fichier de sortie, chaque forme est associée à l'ensemble des métadonnées présentes dans les fichiers du NCA, par le biais d'une fusion entre le fichier des formes et le fichier résumant les métadonnées. L'opération de comparaison entre les deux fichiers est pratiquée par une simple commande incluant l'opérateur %in% de R, et la suite du traitement des données repose principalement sur des commandes du package dplyr (WICKHAM et al. 2023). Ainsi, la commande en (3.8) repère les lemmes de la liste des formes dotées d'un <e> final qui apparaissent aussi la liste des formes sans <e> final, puis la commande en (3.9) crée un objet contenant toutes les formes dotées d'un <e> final correspondant aux lemmes communs repérés, et la commande en (3.10) produit une vision synthétique en comptant le nombre de formes de chacun de ces lemmes. L'opération est pratiquée en miroir à partir de la liste des formes sans <e> final.

- (3.8) `POS_E$Duplicated <- POS_E$lemma %in% POS_nonE$lemma2`
Glose : crée une variable Duplicated dont l'attribut est TRUE pour chaque forme de POS_E dont le lemme apparaît dans POS_nonE.
- (3.9) `POS_E_duplicated <- POS_E %>%
 filter(Duplicated=='TRUE')`
Glose : crée un nouvel objet POS_E_duplicated contenant uniquement les occurrences de POS_E dont la variable Duplicated est associée à la valeur TRUE.
- (3.10) `POS_E_freq <- POS_E_duplicated %>%
 count(lemma, sort=TRUE)`
Glose : crée une table des fréquences des lemmes de POS_E_duplicated.

Les deux tables de fréquence des lemmes à variation <e>/∅, issues de (3.10) et de son équivalent miroir pour les formes sans <e> final, sont ensuite fusionnées par la commande en (3.11). Cette commande génère un objet POS_merged_freq indiquant, pour chaque lemme, le nombre de formes dotées d'un <e> final (n.E) et le nombre de formes sans <e> final (n.nonE), qui est ensuite exporté au format .xlsx. Ce type est fichier est illustré en table 3.1, qui reproduit les dix premières lignes du fichier pour les adjectifs au féminin singulier.

- (3.11) `POS_merged_freq <- merge (POS_E_freq,
 POS_nonE_freq,
 by="lemma")`
Glose : crée un objet POS_merged_freq en fusionnant, par communauté de lemmes, les fréquences des lemmes de POS_E_freq et POS_nonE_freq.

Du point de vue de l'utilisateur. Du point de vue de l'utilisateur, cette première étape se résume à l'exécution du script MASTER_pre_treatment_R_form_SRC.R. Ce script charge les packages nécessaires, demande à l'utilisateur de lui désigner les fichiers de données, via l'ouverture d'un navigateur de fichiers, et lui demande quel est le nom de la variable qu'il veut utiliser (qui vient remplacer la mention [POS] dans le nom des fichiers). Le script fait ensuite appel à l'un des trois scripts de pré-traitement disponibles, selon que l'utilisateur veut travailler sur

WICKHAM et al. (2023). *dplyr : A Grammar of Data Manipulation.*

2: En langage R, « § » permet de sélectionner une colonne au sein d'un jeu de données tabulaires; ainsi POS_E\$lemma renvoie les valeurs de la colonne lemma du jeu de données POS_E. Nous ne commentons pas les opérateurs ou commandes dont le sens est plus ou moins transparent, comme « <- », « == », filter ou count. Dans la syntaxe du package dplyr, l'opérateur %>% sert à créer un flux d'opérations (en. pipe), équivalent au symbole « + » de la syntaxe native de R.

TABLE 3.1 – Reproduction des dix premières lignes de R_ADJ_femi_sg_merged_freq.xlsx.

lemma	n.E	n.nonE
agu	20	3
autre	2505	15
autretel	6	5
bas	72	3
bel	135	7
benëoit	61	1
bis	27	5
blanc	191	3
bon	1645	56

l'ensemble des formes contenues dans ses fichiers d'*input*, ou qu'il veut les filtrer par genre et par nombre, ou par genre, par nombre et par cas. De cette manière, le risque d'erreur de manipulation est réduit.

Intervention de l'utilisateur sur [POS]_merged_freq. Cette étape donne lieu, avant la troisième étape, à deux occasions, pour l'utilisateur, d'intervenir dans les données pour exclure des formes non pertinentes. La première consiste à éditer directement le fichier [POS]_merged_freq.xlsx, ce que nous avons fait en ajoutant manuellement deux variables pour chaque lemme : la première (*Var*) qualifie sous forme binaire (Oui/Non) chaque lemme détecté par Comparalem selon que la variation en question correspond bien au but de la requête CQP, la seconde (*Var_stringent*) ajoute un critère, selon que le schwa final est étymologique ou non. Cela permet d'exclure les lemmes qui n'ont pas leur place dans ce tableau de données. Ainsi, dans les dix premières lignes de R ADJ_femi_sg_merged_freq.xlsx reproduites en table 3.1, figurent des lemmes qui ne témoignent pas d'une réelle alternance <e>/∅ qui soit phonologique. Les trois occurrences du lemme « *agu* » sans <e> qualifient le mot *glaiive*, auquel le DMF donne le genre féminin mais auquel le GD donne le genre masculin. Ici, l'hypothèse la plus probable est que le *glaiive* auquel se rapportent les trois occurrences *agu* est au masculin et que, donc, c'est l'annotation du genre qui est trompeuse. La variable *var* est donc éditée avec la valeur « Non », pour exclure ce lemme, puisqu'il ne dispose pas, au féminin singulier, de formes sans <e> final. De même, les 15 formes sans <e> final du lemme « autre » correspondent à des formes *autra* qu'on trouve dans des textes de la région Vendée, Deux-Sèvres ; comme ces <a> pour schwa final sont courants dans cette région, il ne s'agit pas ici d'une variation <e>/∅, mais d'une variation <e>/<a> pour schwa — ainsi, Comparalem permet également la détection de cette variation.

Un autre exemple d'exclusion nécessaire est le lemme *autretel* ; descendant de l'adjectif latin de deuxième classe *TALIS*, cet adjectif est naturellement épïcène et le schwa final n'y est pas étymologique ; ici, il y a bien une variation <e>/∅ qui ne provient pas du genre, mais provient de l'analogie entre ces formes étymologiquement épïcènes et les formes fléchies en genre, qui aboutit à construire un féminin *autretele* sur le fondement d'une base *autretel-* à laquelle s'ajoute un suffixe de féminin *-e* (au lieu d'un suffixe zéro épïcène). Cette fois-ci, c'est la variable *Var_stringent* qui prend la valeur « Non », puisqu'elle détermine si le schwa attendu est motivé par l'étymologie.³ En cas d'attestation d'un nom dont le genre fluctue (ici le nom auquel se rapporte l'adjectif), le lemme a été exclu puisqu'on ne peut attester d'une variation phonologique <e>/∅ si l'on ne sait pas quel est le genre du mot. Nous reproduisons en table 3.2 cette étape de tri dans les données ; ne sont conservés en définitive que les lemmes dont la variable *Var_stringent* est « Oui ».

La vérification des occurrences peut s'effectuer soit dans TXM (en rédigeant une requête CQP *ad hoc*), soit dans l'environnement de développement dans lequel l'utilisateur exécute R ; dans R Studio, le code déclenche l'affichage de tables contenant les formes, leurs attributs et leurs références. La deuxième étape du traitement des données par Comparalem fait référence à ces lemmes qui disposent de la valeur « Oui » pour la variable *Var_stringent*. En inscrivant une valeur « Non » pour l'une

3: L'annotation de ces données est une tâche fort chronophage, aussi il arrive que nous ayons édité directement la variable *Var* au lieu d'éditer la variable *Var_stringent* dans des cas analogues à celui-ci ; cela est sans effet sur la qualité des données, dans la mesure où *Var=="Non"* implique systématiquement *Var_stringent=="Non"*.

TABLE 3.2 – Reproduction des dix premières lignes de R ADJ_femi_sg_merged_freq.xlsx, avec tri des lemmes.

lemma	n.E	n.nonE	Var	Var_stringent	Remarques
agu	20	3	Non	Non	Toujours avec <i>glaive</i> , qui est masc. (GD; DMF donne fém.)
autre	2505	15	Non	Non	Formes <i>autra</i> (<i>alia</i> , <i>poit</i>)
autretel	6	5	Oui	Non	Épicène
bas	72	3	Non	Non	Une occ. devant <i>gent</i> (masc. ou fém.), et 2 potentiellement adverbiales (<i>ne dei plus bas femme choisir</i>)
bel	135	7	Non	Non	6 occ. <i>bela</i> (<i>poit</i>); 1 occ. <i>beu sire</i> mal annotée en genre
benëoit	61	1	Oui	Oui	et <i>aigue beneoit boive</i>
bis	27	5	Non	Non	Toujours dans <i>porpre bis</i> , adj. de couleur que le DMF donne uniquement au masc.
blanc	191	3	Oui	Oui	1 occ. OK (<i>blanc ad la barbe</i> , <i>rolandox</i>), 2 occ. avec <i>chainse</i> (« chemise »), masc. ou fém. (DMF)
bon	1645	56	Oui	Oui	5 occ. <i>bona</i> (<i>poit</i>), 4 occ. devant <i>œuvre</i> (masc. ou fém. : anti x2, moral2 x2), 6 occ. masc. ou en compo. (<i>de bon aire</i> : auc; aucchants; fablesB; fablesC; fablesY; ipo), 1 occ. adverbiale (<i>bun est a achateir</i> : darm), 4 occ. avec <i>amur</i> (masc. ou fém. : elid; ipo x2; ndchar), 3 occ. masc. (<i>de bon servoise</i> (<i>ipo</i>); <i>a son servoise était mut bon</i> (<i>ipo</i>); <i>soient buens soient mals</i> (<i>moral2</i>))

des variables Var ou Var_stringent, l'utilisateur peut ainsi exclure des lemmes de la suite de l'analyse.

Intervention de l'utilisateur : [POS]_irrelevant forms. Néanmoins, exclure un lemme n'est valable que lorsque l'ensemble des occurrences du lemme doivent être exclues, ce qui n'est pas toujours le cas. En table 3.2, les lemmes « blanc » et « bon » disposent de formes qu'il faut retrancher, mais aussi de formes qu'il faut conserver. Pour « blanc », il faut conserver l'occurrence de *rolandox* mais exclure les deux occurrences où *blanc* se rapporte à *chemise*, qui peut être masculin en ancien et moyen français.⁴ Pour *bon*, il faut exclure les 23 formes commentées en table 3.2, mais les 33 autres formes doivent être conservées. Pour ce faire, l'utilisateur peut éditer un fichier tabulaire, que nous dénommons conventionnellement [POS]_irrelevant forms mais dont le nom est libre, dans lequel il inscrit la forme, le lemme, l'identifiant unique du mot au sein du texte et l'identifiant du texte. Ce fichier sera appelé lors de la seconde étape de d'exécution de Comparalem, et permettra ainsi l'exclusion des occurrences en question. Les 10 premières lignes de ce fichier pour les adjectifs féminins singuliers sont reproduites en table 3.3.

rolandox : ANON. (1100*). *La Chanson de Roland*. Ms. O; Dees : Angleterre; compo. N.-O., ca 1100; ms. a.n., ca 1137; CR1 (SEGRE 1971); DEAF : Ro1S (ici p. 88).

4: **guigP**, p. 8 : *bel drap de sainsil blanc*; **lanvalP**, p. 11 : *de cainse blanc et de cemise de tout li coste li paroiert*.

guigP : MARIE DE FRANCE (1165*). *Le Lai de Guigemar*. Ms. P; Dees : nil; compo. N.-O., ca 1165; ms. pic., ca 1290; nil (transcr. inédite sans transcripteur indiqué).

lanvalP : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. P; Dees : nil; compo. N.-O., ca 1165; ms. pic., ca 1290; MS2 (MARIE DE FRANCE 1958); DEAF : MarieLanvR².

3.2.1.3 Seconde étape : observer la distribution des variantes

La seconde étape d'utilisation de Comparalem est celle qui représente la plus grande complexité au niveau du code mais, une nouvelle fois, elle demeure assez simple dans la perspective de l'utilisateur. À cette étape, le programme dispose de deux comportements différents, selon que l'analyse qui lui est demandée est contextuelle (devant #V versus

TABLE 3.3 – Reproduction des dix premières lignes du fichier d'exclusion ADJ_irrelevant forms.xlsx concernant les adjectifs féminins singuliers.

word	lemma	ref1	id
blanc	blanc	s170993	guigP
blanc	blanc	s205165	lanvalP
bona	bon	s326281	poit
bona	bon	s326803	poit
bona	bon	s326980	poit
bona	bon	s329457	poit
bona	bon	s329674	poit
bon	bon	s21105	anti
bon	bon	s21107	anti
bon	bon	s28351	auc
bon	bon	s28706	aucchants

devant #C) ou non. L'analyse non contextuelle n'est que peu utilisée dans notre travail, aussi, dans ce qui suit, nous détaillons le comportement du script et de ses étapes dans la mise en place d'une analyse contextuelle. Les informations nécessaires à la compréhension de l'analyse non contextuelles seront abordées rapidement lorsqu'une analyse non contextuelle sera utilisée.

La seconde étape de Comparalem prend comme *input* quatre fichiers de données, deux fichiers de filtre, et la table des métadonnées du NCA. Ces quatre fichiers correspondent à la liste des formes sans <e> et avec <e>, devant initiale consonantique et devant initiale vocalique. Sur ces fichiers, le programme applique une série d'opérations, réparties dans différents scripts. Ces étapes divergent selon que l'analyse s'opère sur l'ensemble des données fournies en *input*, ou selon que des tris sont effectués parmi ces données. Ces étapes, que nous abordons brièvement par la suite, sont résumées en (3.12). Les étapes de filtrage des données sont optionnelles, soit parce qu'elles ne modifient pas le nom des variables, soit parce qu'un script de contournement est disponible.

(3.12) Étapes du deuxième tour de Comparalem

- (a) Import des données;
- (b) Filtrage en fonction des lemmes;
- (c) Filtrage en fonction des formes;
- (d) Filtrage en fonction des dates;
- (e) Filtrage en fonction du genre et du nombre ou du genre, du nombre et du cas;
- (f) Production de sorties graphiques diachroniques générales;
- (g) Production de sorties graphiques par lemmes;
- (h) Production de données quantitatives;
- (i) Production d'une sortie cartographique synchronique;
- (j) Production de sorties cartographiques diachroniques;
- (k) Export des sorties graphiques et quantitatives.

Import des données. Lors de cette étape, le programme demande à l'utilisateur, *via* l'ouverture d'un navigateur de fichier, de lui indiquer la

position des fichiers de données issues des requêtes CQP dans TXM. Dans le cas d'une analyse contextuelle, quatre fichiers sont demandés (formes avec <e> final et formes sans <e> final, devant initiale consonantique et devant initiale vocalique); les fichiers sont demandés dans l'ordre reproduit en (3.13). L'import des fichiers dans l'ordre représente la principale source d'erreur potentielle dans l'utilisation du programme, dans la mesure où l'ouverture du navigateur de fichier par le script ne prévoit pas l'énoncé du type de fichier demandé, mais ces erreurs se repèrent aisément lors du traitement des données par la suite. À la suite de l'import des fichiers, le script effectue un certain nombre d'opérations de transformation des données, dont l'association des métadonnées du texte à chaque forme, à partir de l'import (automatique) du fichier de métadonnées du NCA.

- (3.13) Analyse contextuelle, ordre d'importation des fichiers
- (a) POS_E_C : fichier correspondant à la requête pour les mots se terminant par <e> pour /ə/ devant initiale consonantique;
 - (b) POS_E_V : fichier correspondant à la requête pour les mots se terminant par <e> pour /ə/ devant initiale vocalique;
 - (c) POS_nonE_C : fichier correspondant à la requête pour les mots ne se terminant pas par <e> pour /ə/, situés devant initiale consonantique;
 - (d) POS_nonE_V : fichier correspondant à la requête pour les mots ne se terminant pas par <e> pour /ə/, situés devant initiale vocalique.

Filtrage en fonction des lemmes. Cette étape est facultative, et peut être contournée par exemple dans l'optique où l'utilisateur voudrait ne travailler que sur un seul lemme. Si elle est déclenchée, elle demande à l'utilisateur, *via* l'ouverture d'un navigateur de fichier, de lui indiquer la position du fichier [POS]_merged_freq contenant les lemmes sujets à variation <e>/Ø. Cette étape procède à l'effacement des données de tous les lemmes qui ne sont pas représentés dans ce fichier [POS]_merged_freq avec la valeur « Oui » pour la variable « Var_stringent »

Filtrage en fonction des occurrences exclues. Cette étape est facultative, et peut être contournée par exemple dans l'optique où l'utilisateur n'a pas de formes à exclure. Si elle est déclenchée, elle demande à l'utilisateur, *via* l'ouverture d'un navigateur de fichier, de lui indiquer la position du fichier [POS]_irrelevant_forms contenant les formes à exclure ponctuellement ainsi que leurs références. Cette étape procède à l'effacement des données de toutes les occurrences contenues dans le fichier [POS]_irrelevant_forms.

Filtrage en fonction des dates. Cette étape est facultative, et peut être contournée par exemple si l'utilisateur veut travailler sur l'ensemble de son corpus. Dans notre cas, elle a toujours été appliquée, de manière à exclure les textes postérieurs à 1322 (cf. p. 38). Le script correspondant à cette restriction temporelle a été rédigé spécialement pour cette exclusion, mais peut aisément être réécrit par un utilisateur qui connaîtrait le

langage R. Cette étape supprime des données toutes les formes datées après 1322.

Filtrage en fonction de la flexion. Cette étape est facultative, et peut-être contournée par exemple si l'utilisateur travaille sur une partie du discours dans son ensemble et non sur un sous-ensemble fléchi de formes de cette partie du discours. Si elle est déclenchée, elle utilise un *input* que l'utilisateur a fourni dans la console R pour définir quelle est la sous-catégorie qu'elle va conserver. La saisie de l'utilisateur prend la forme « genre_nombre » ou « genre_nombre_cas » (p.ex. femi_sg ou femi_sg_suj), combinant par là les valeurs de deux ou trois variables différentes. Cette étape supprime des données toutes les occurrences qui ne cumulent pas les deux ou trois propriétés saisies par l'utilisateur.

Production de sorties graphiques diachroniques. Cette étape effectue un certain nombre d'opérations sur les données, et produit des sorties graphiques diachroniques. Ces sorties graphiques représentent le taux de formes sans <e> final par texte pour les lemmes à variation <e>/Ø, ainsi que quelques mesures quantitatives et une mesure, par le test de Student, de l'influence du contexte subséquent sur la variation <e>/Ø. Cette sortie graphique peut prendre deux formes principales, selon qu'elle présente le taux de <e> en fonction du contexte subséquent, ou qu'elle présente le taux d'initiales vocaliques en fonction de la présence ou de l'absence d'un <e> final. Nous utilisons principalement la première sortie, illustrée en fig. 3.2 pour les adjectifs féminins singuliers. La seconde sortie est illustrée en fig. 3.3, toujours pour les adjectifs féminins singuliers, nous n'en faisons pas grand usage dans la rédaction de cette thèse, mais elle a été utilisée pour le contrôle du programme et des occurrences.

Production de sorties graphiques par lemme. Cette étape produit deux sorties graphiques qui permettent de visualiser non plus les taux globaux des variantes Ø par texte, mais le taux de variantes Ø lemme par lemme. Selon le nombre d'occurrences et de lemmes, les graphiques produits sont plus ou moins lisibles, mais ils aident à avoir un aperçu de la façon dont les différents lemmes sont atteints par l'alternance <e>/Ø dans le temps. Deux graphiques sont produits, selon que les occurrences se trouvent devant une initiale vocalique ou consonantique, comme illustré en fig. 5.2 et 5.3. Les graphiques s'accompagnent d'un extrait des données affichant les fréquences des dix lemmes disposant du plus de variantes Ø.

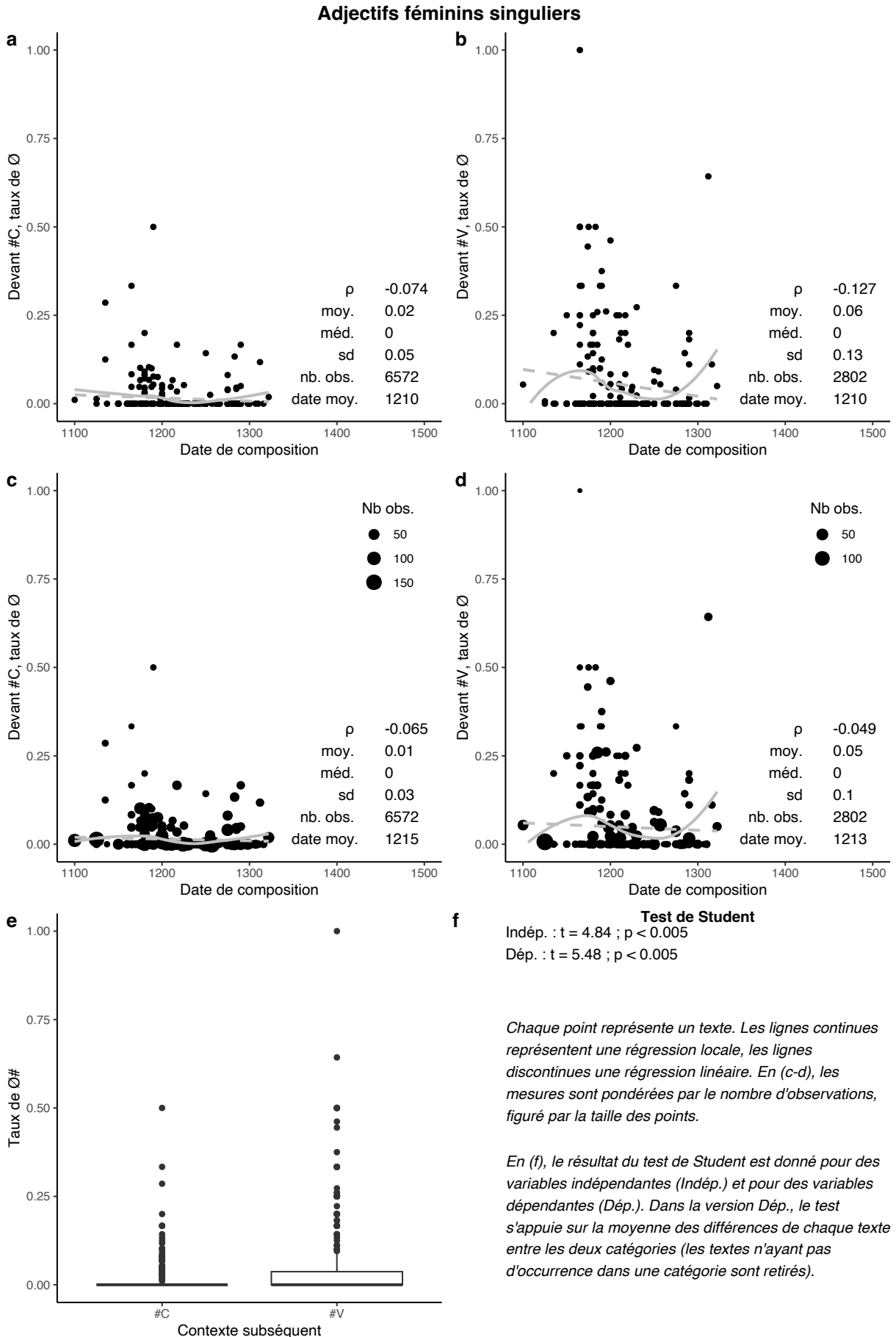


FIGURE 3.2 – Distribution diachronique de la variation finale <e>/Ø en fonction du contexte subséquent, adjectifs féminins singuliers.

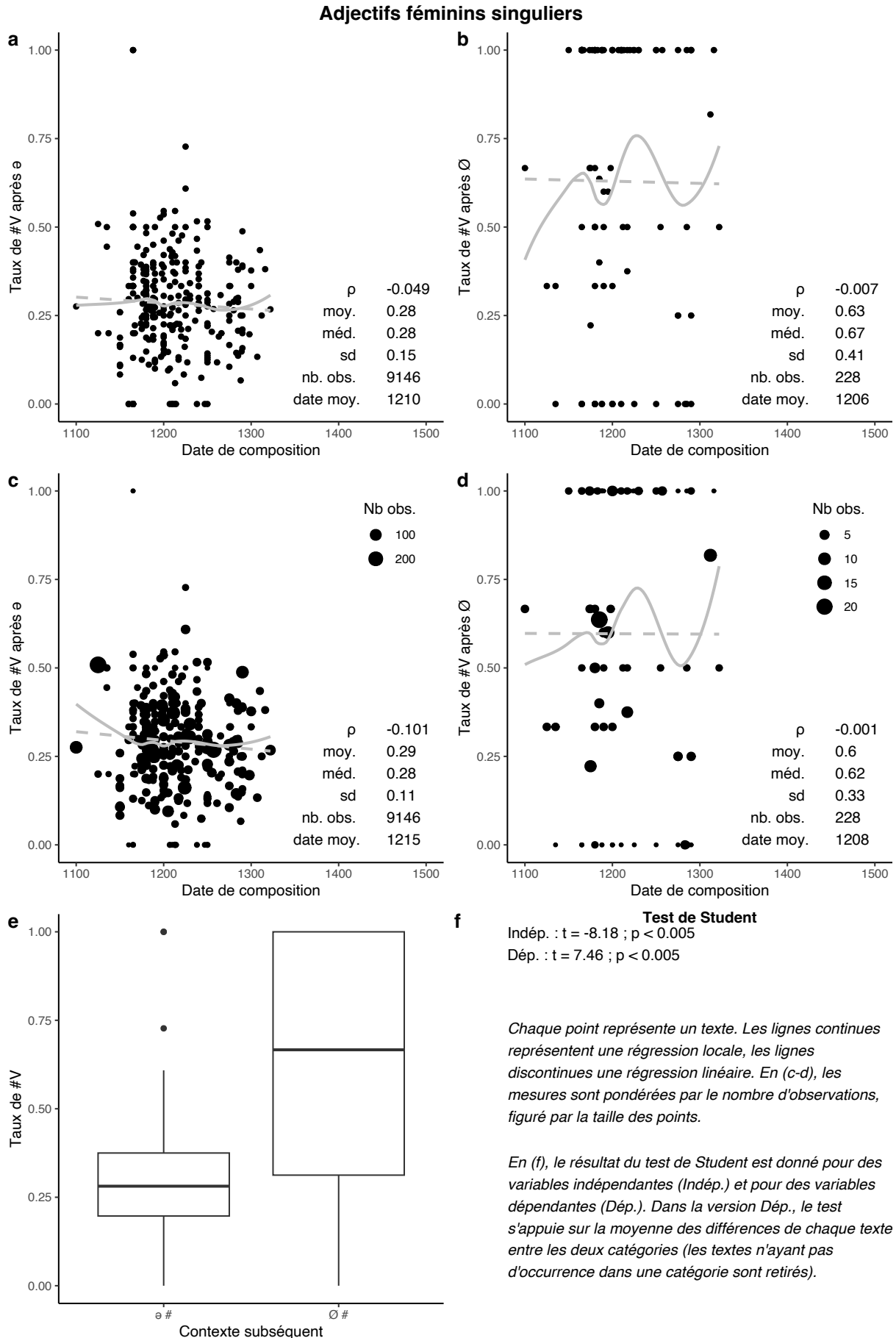
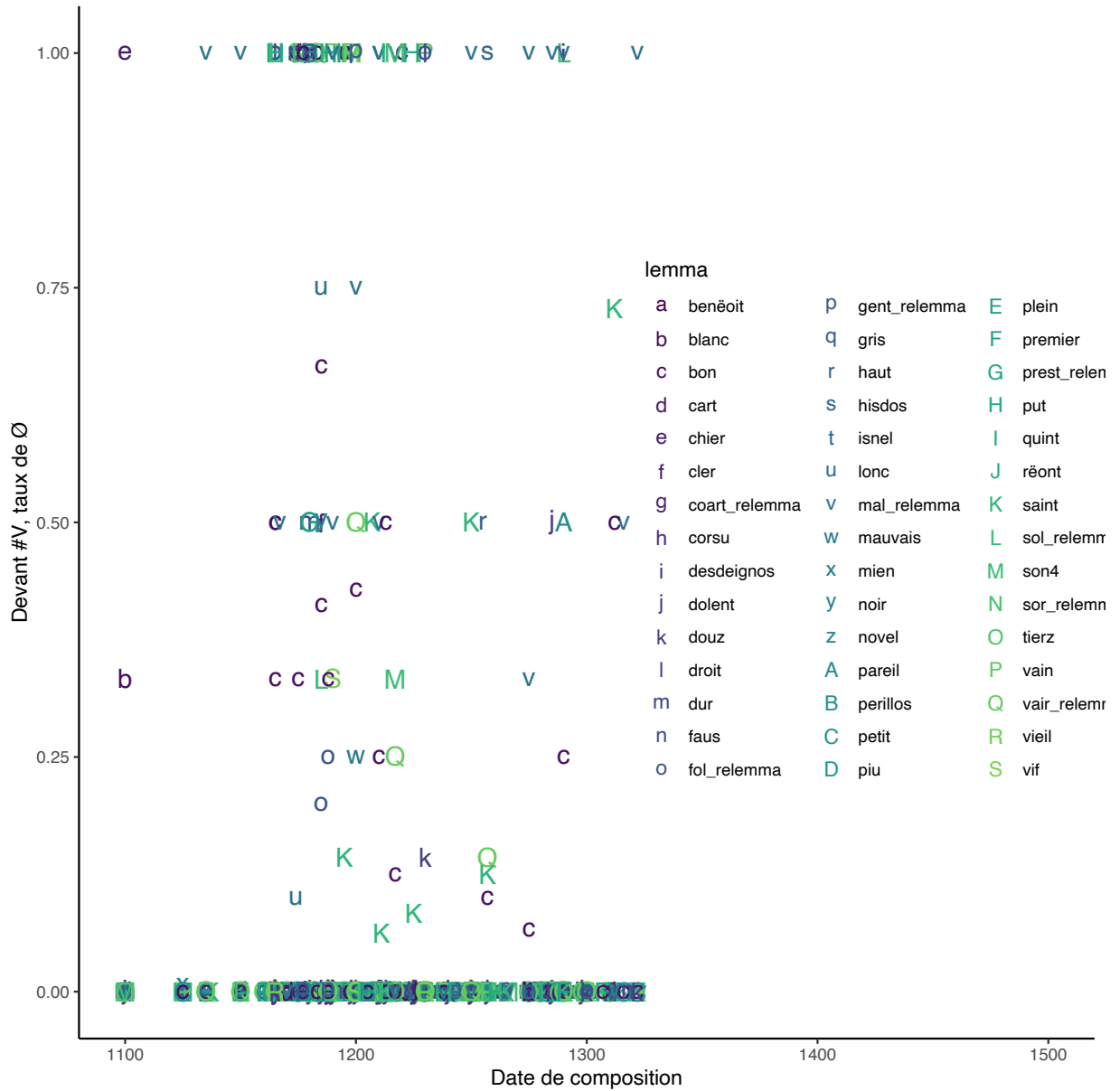
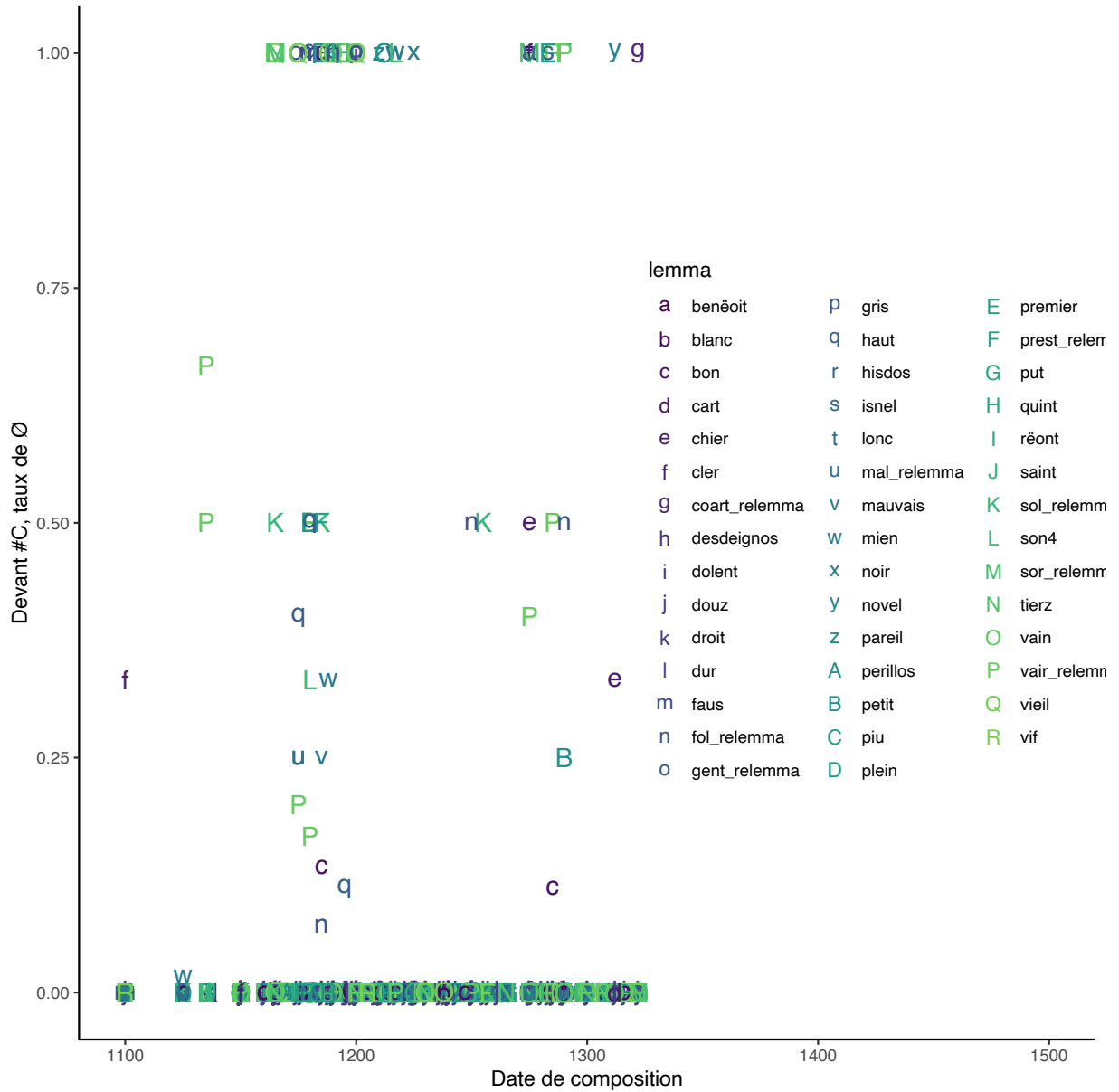


FIGURE 3.3 – Distribution diachronique des initiales consonantiques ou vocaliques en fonction de la présence ou absence d'un <e> en fin du mot précédent, adjectifs féminins singuliers.



lemma	nb.e	nb.∅	tot.	tx.∅
desdeignos	1	1	2	0.50
corsu	4	3	7	0.43
pareil	2	1	3	0.33
droit	6	2	8	0.25
mal_relemma	96	26	122	0.21
put	23	6	29	0.21
quint	7	1	8	0.12
sol_relemma	55	7	62	0.11
son4	35	4	39	0.10
hisdos	10	1	11	0.09
[autres : 35]	2427	84	2511	0.03

FIGURE 3.4 – Distribution diachronique des alternances <e>/∅ selon les lemmes, adjectifs féminins singuliers suivis d’une initiale vocalique.



lemma	nb.e	nb.Ø	tot.	tx.Ø
gris	4	2	6	0.33
sor_relemma	10	2	12	0.17
coart_relemma	5	1	6	0.17
piu	5	1	6	0.17
isnel	7	1	8	0.12
prest_relemma	28	2	30	0.07
faus	61	3	64	0.05
pareil	23	1	24	0.04
vieil	213	9	222	0.04
son4	124	5	129	0.04
[autres : 34]	6000	65	6065	0.01

FIGURE 3.5 – Distribution diachronique des alternances <e>/<Ø> selon les lemmes, adjectifs féminins singuliers suivis d’une initiale consonantique.

Production de données quantitatives. Cette étape permet la génération de deux tables synchroniques récapitulatives. La première représente la distribution non contextuelle de l’alternance <e>/Ø dans les données étudiées, tandis que la seconde représente cette distribution en fonction du contexte. Ces tableaux sont reproduits en tables 3.4 et 3.5. En table 3.4, les colonnes « Nb.var » et « Taux » représentent le nombre de lemmes qui sont sujets à l’alternance <e>/Ø et le nombre d’occurrences totales de ces lemmes, tandis que les colonnes « Nb.Var.Ø » et « Taux.Ø » représentent le nombre d’occurrences de ces lemmes sans <e> final. En table 3.5, les colonnes « X_C » et « X_V » représentent respectivement les contextes pré-consonantiques et pré-vocaliques, la ligne « Voyelle » contient le nombre d’occurrences se terminant par un <e> final (ou une autre voyelle dans le cas des monosyllabes fonctionnels) et la ligne « Apocope » représente le nombre d’occurrences sans <e> (ou autre voyelle dans le cas de certains monosyllabes fonctionnels) final.

TABLE 3.4 – Données récapitulatives non contextuelles pour les adjectifs féminins singuliers.

ADJ_femi_sg	Nb.	Nb.var.	Taux	Nb.Var.Ø	Taux.Ø
Lemmes	1756	48	2.73 %		
Occ.	130298	9428	7.24 %	261	0.2 %

TABLE 3.5 – Données récapitulatives contextuelles pour les adjectifs féminins singuliers.

ADJ_femi_sg	X_C	X_V
Voyelle	6494	2673
Apocope	117	144
Taux apo.	1.77 %	5.11 %
Somme	6611	2817

Production de sorties cartographiques synchroniques. À partir des données retravaillées à l’issue des différents filtres, Comparalem produit des cartes du domaine d’oïl reproduisant, grâce aux *shapefiles* établis par Guylaine Brun-Trigaud pour le projet ADE22 (SCHEER et al. 2022), les régions de Dees. Ces cartes indiquent la moyenne des taux de variantes Ø des textes de chaque région ; il ne s’agit donc pas de la moyenne des taux de variantes Ø des régions après mélange des différents textes, mais bien de la moyenne des taux de chaque texte, région par région. En effet, étant donné la variation très forte entre les textes sur certaines requêtes et les différences importantes dans le nombre d’occurrences de chaque texte, il nous a semblé plus approprié de considérer la moyenne des textes plutôt que la moyenne des occurrences, de manière à ne pas déséquilibrer une région qui aurait, par exemple, un seul texte possédant beaucoup plus d’occurrences Ø que tous les autres textes. De ce fait, la représentation des régions conserve une certaine intégrité, en nivelant l’effet des *outliers*, des textes ayant des résultats extrêmes.

Les cartes sont accompagnées de mesures sommaires, reflétant les taux régionaux extrêmes, la moyenne des taux régionaux et l’écart type (σ) de leur distribution. Dans une analyse contextuelle, cette étape produit trois cartes : la première représente les taux de Ø devant initiale consonantique (fig. 3.6 a), la deuxième devant initiale vocalique (fig. 3.6 b), et la troisième représente ce que nous appelons le taux de prévalence de l’élision

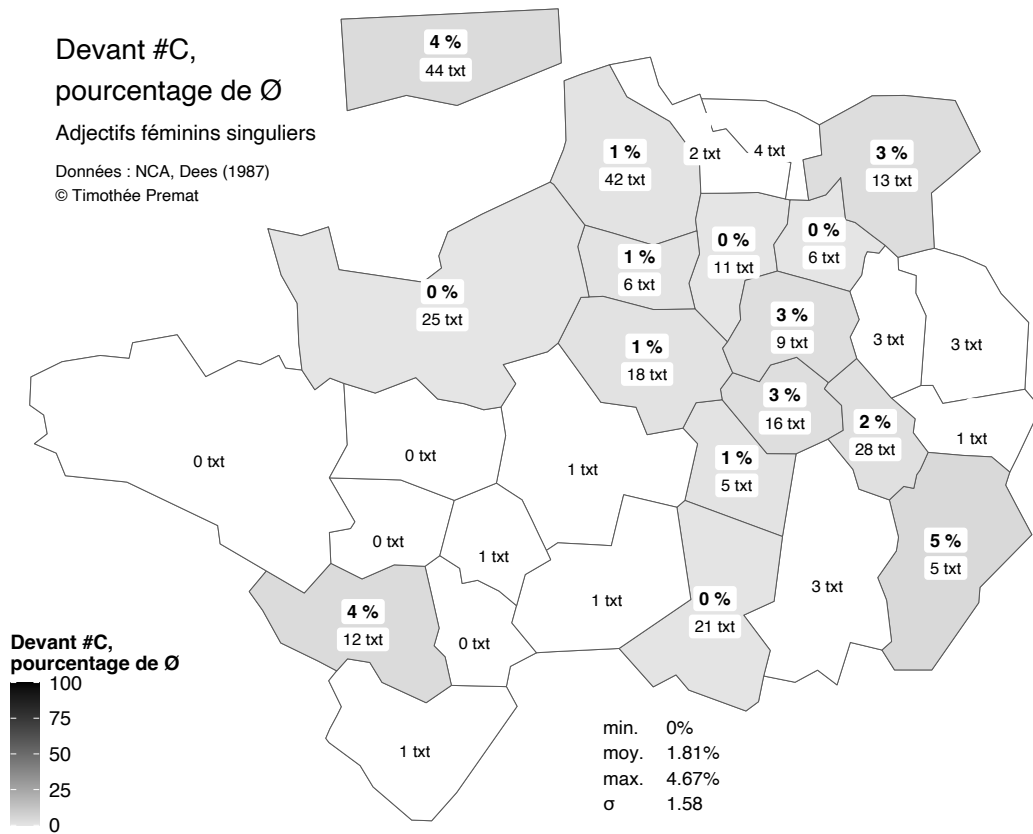
(fig. 3.7), qui correspond à la soustraction du taux de Ø devant #C au taux de Ø devant #V (3.14) et qui se présente comme une mesure rapide de l'influence du contexte sur les alternances <e>/Ø, allant de -1 (Ø n'apparaît que devant #C) à +1 (Ø n'apparaît que devant #V).

$$(3.14) \quad \text{Prévalence de l'élision} = \text{taux de } \emptyset \text{ devant } \#V - \text{taux de } \emptyset \text{ devant } \#C$$

Sur ces cartes, nous avons défini un seuil arbitraire en-deçà duquel une région n'est pas considérée suffisamment solide pour être représentative; ce seuil est fixé à cinq textes. Les deux premières cartes sont en nuances de gris, le blanc étant réservé aux régions neutralisées par manque de textes; la troisième est sur un dégradé du rouge au vert, le blanc correspondant à la fois aux régions neutralisées par manque de textes et aux régions dont le taux de prévalence de l'élision est nul. Enfin, ces cartes portent sur l'ensemble de la période considérée, qu'elles écrasent dans une synchronie artificielle, à la manière des cartes de l'*Atlas* de DEES (1987).

Production de sorties cartographiques diachroniques. La seconde étape cartographique propose un découpage périodique des textes, selon des bornes qui peuvent être définies par l'utilisateur. Par défaut, ce découpage se fait en quatre parties, correspondant aux demi-siècles déjà utilisés en sec. 2.1.5 pour l'illustration de la répartition du corpus; nous avons principalement utilisé ces bornes par défaut de manière à permettre une comparaison aisée entre les différentes données analysées, à l'exception des deux sections dédiées aux schwas internes (chap. 11), où les données dont nous disposons pour la syncope et la réduction des hiatus internes ne nous permettaient pas un découpage aussi fin. Dans la mesure où les dates ne sont jamais exactes et où la distribution diachronique du corpus affaiblit le premier demi-siècle et renforce artificiellement le dernier 'demi-siècle' (qui fait 72 ans au lieu de 50, allant de 1250 à 1322), nous avons choisi de faire basculer les années charnières dans le demi-siècle précédant : ainsi, l'année 1150 appartient au premier demi-siècle (1100-1150) et non au second (1151-1200), etc. - cette manipulation est sans gravité dans la mesure où une écrasante majorité des dates résulte d'une normalisation de dates approximatives. De manière à ne pas multiplier inutilement le nombre de cartes, cette étape ne produit pas automatiquement les cartes représentant la distribution de Ø devant #C, mais elle produit les cartes représentant la distribution de Ø devant #V (fig. 3.8 et 3.9) ainsi que les cartes de prévalence de l'élision (fig. 3.10 et 3.11); la distribution de Ø devant #C est donc restituable par simple comparaison de la carte de Ø devant #V et de la carte de prévalence de l'élision pour une période donnée. Enfin, étant donné que le nombre de textes par carte est réduit par la division en périodes, le seuil de représentativité d'une région (5 textes par région) est ici aboli.

a



b

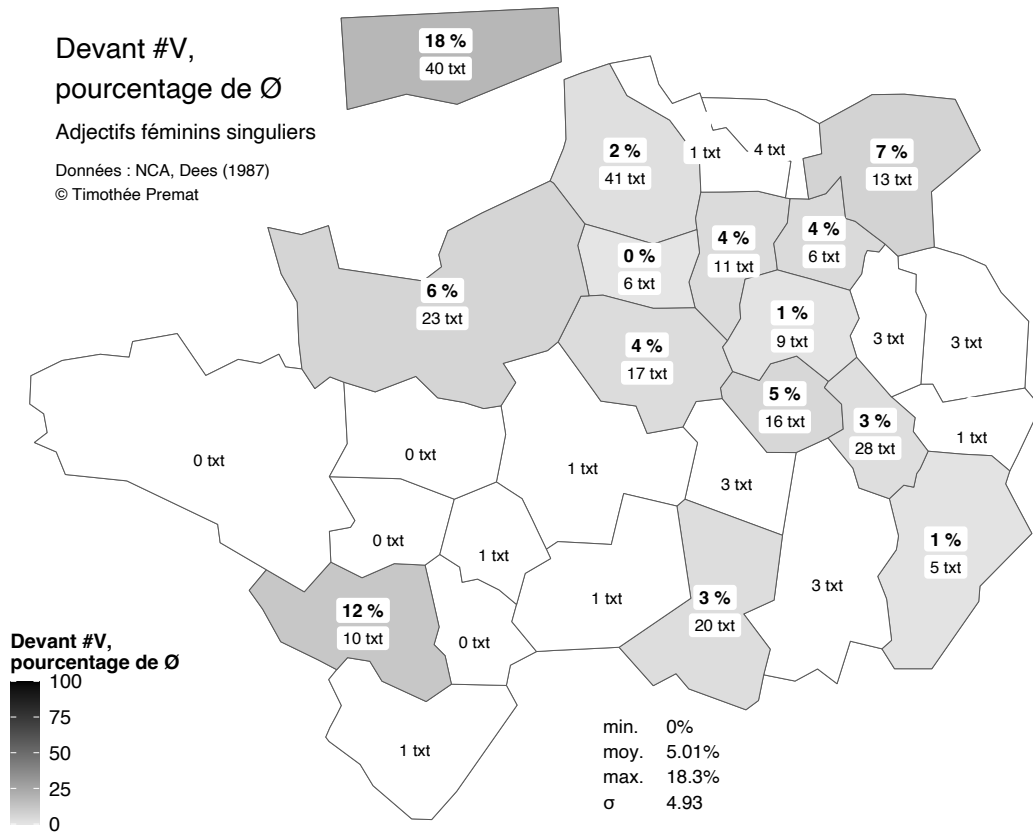
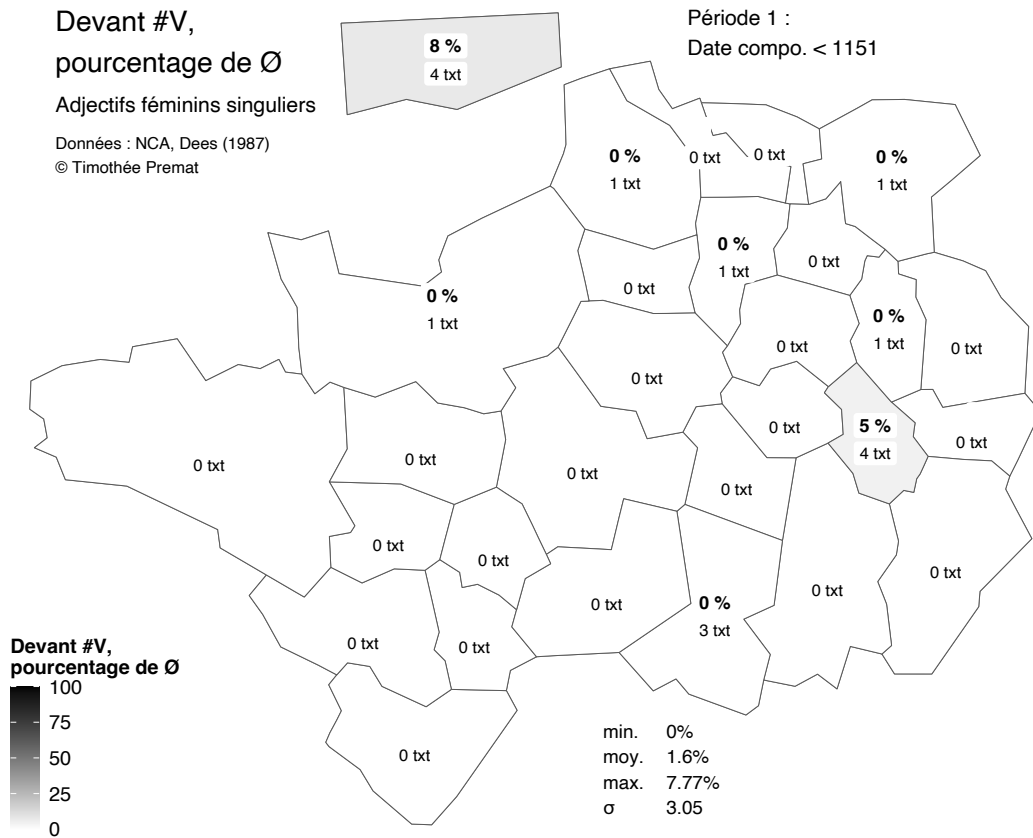


FIGURE 3.6 – Distribution diatopique des alternances <e>/Ø, sur l’ensemble de la période considérée, adjectifs féminins singuliers suivis d’une initiale consonantique (a) et suivis d’une initiale vocalique (b).

a



b

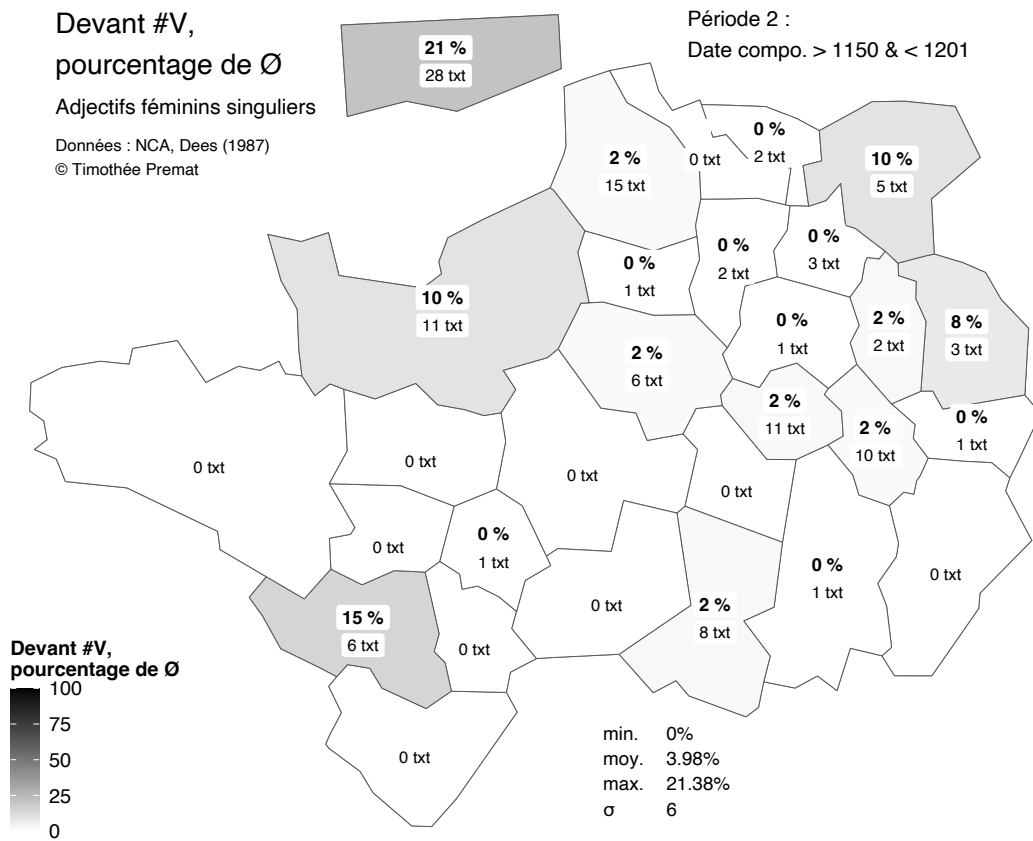
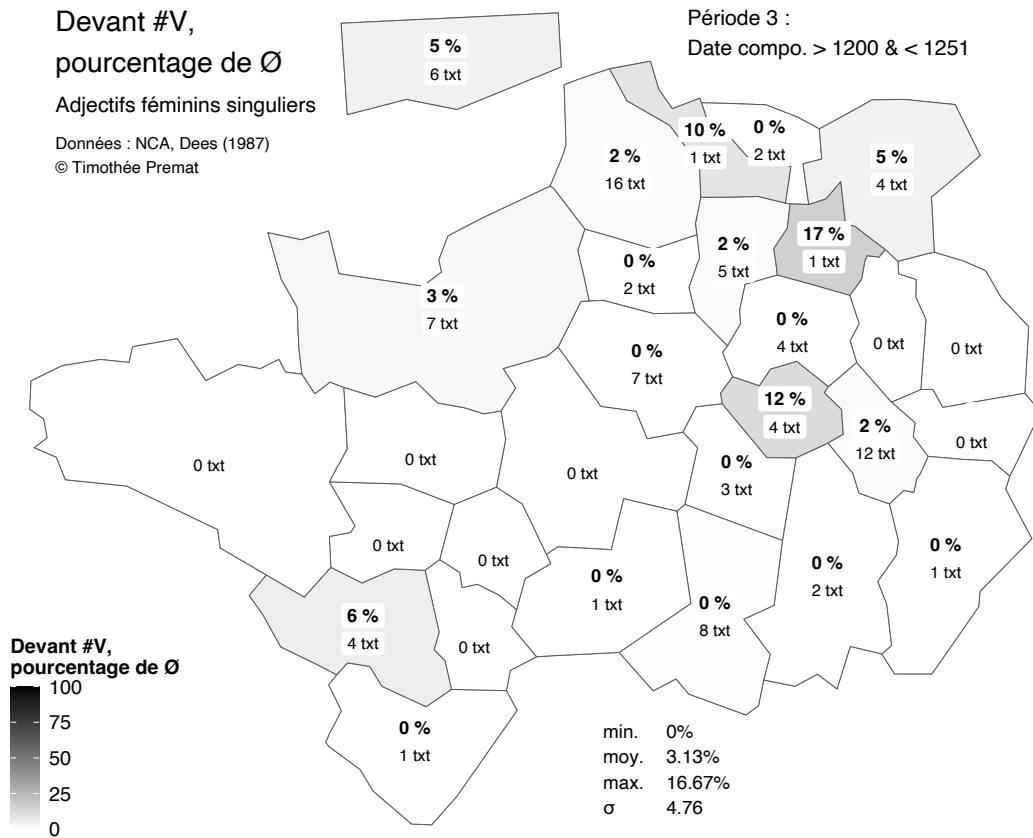


FIGURE 3.8 – Distribution diatopique des alternances <e>/Ø, périodes 1 (1100-1150, a) et 2 (1151-1200), adjectifs féminins singuliers suivis d’une initiale vocalique.

a



b

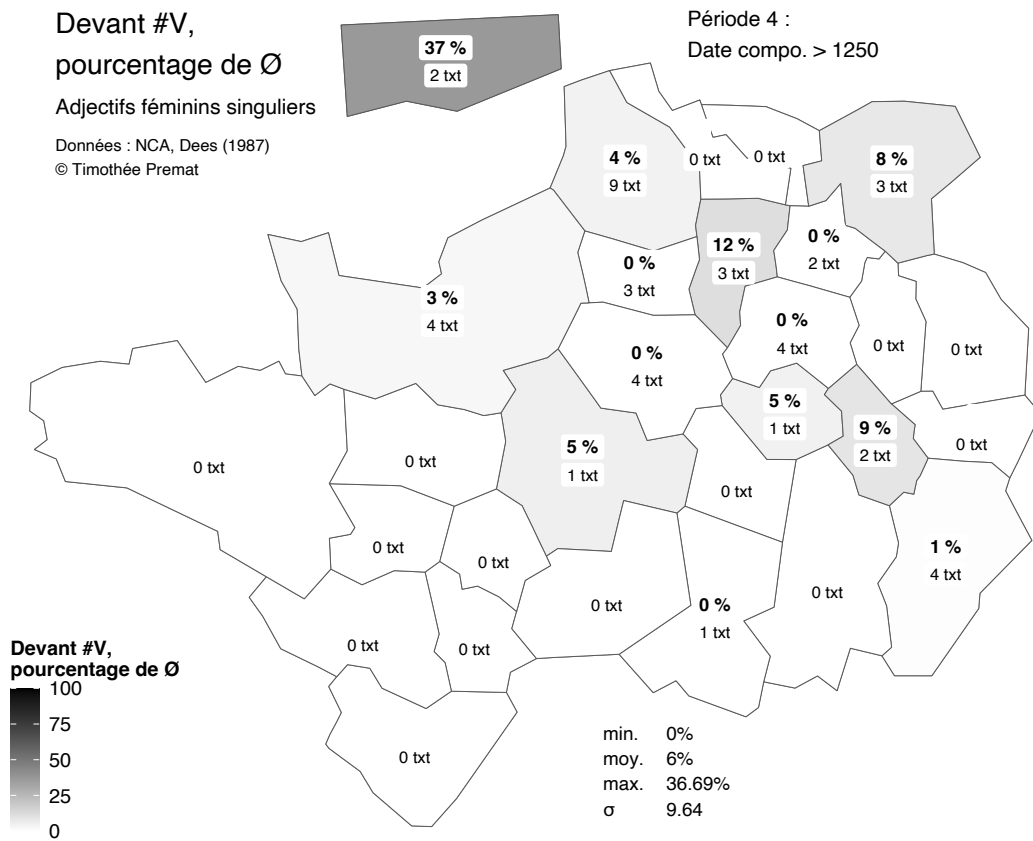
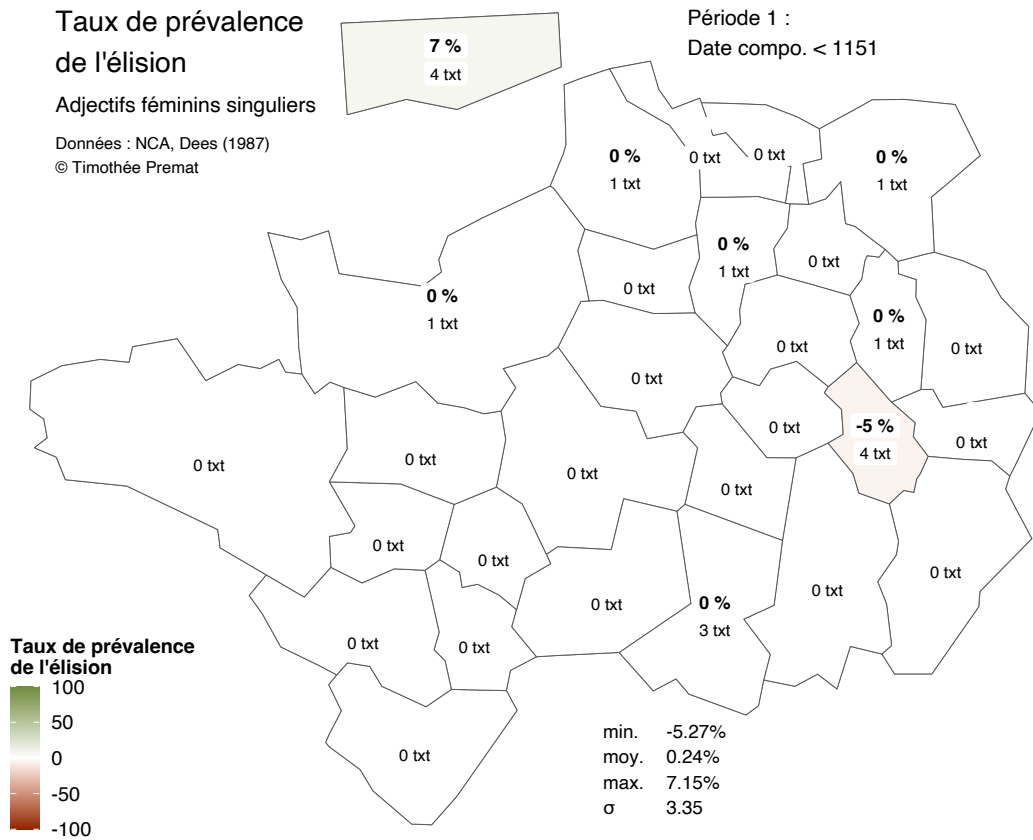


FIGURE 3.9 – Distribution diatopique des alternances $\langle e \rangle / \emptyset$, périodes 3 (1201-1250, a) et 4 (1251-1322), adjectifs féminins singuliers suivis d’une initiale vocalique.

a



b

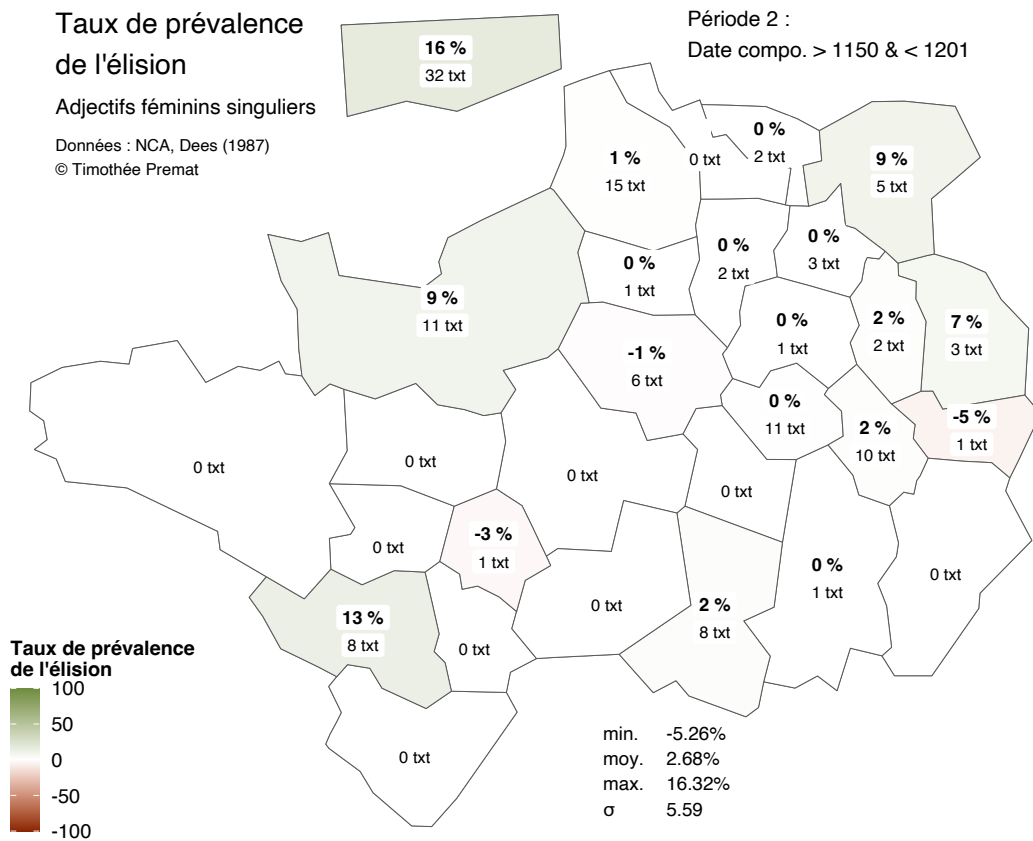
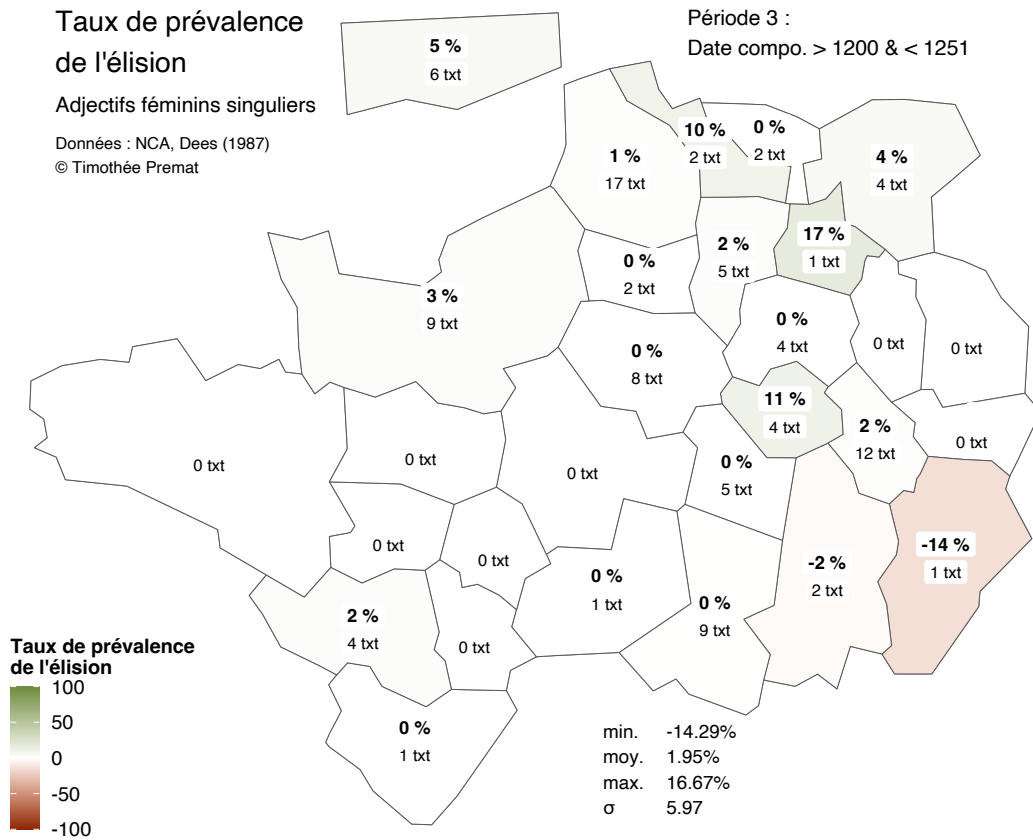


FIGURE 3.10 – Distribution diatopique de la prévalence de l'élision, périodes 1 (1100-1150, a) et 2 (1151-1200), adjectifs féminins singuliers.

a



b

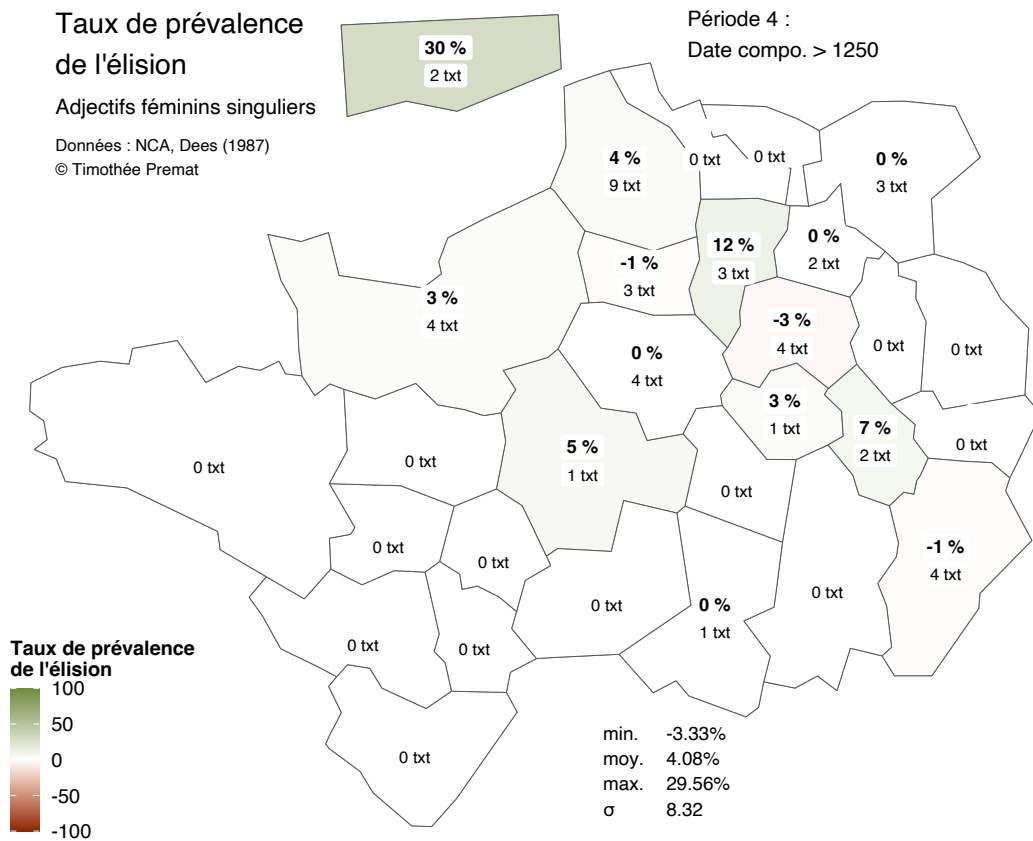


FIGURE 3.11 – Distribution diatopique de la prévalence de l'élision, périodes 3 (1201-1250, a) et 4 (1251-1322), adjectifs féminins singuliers.

Export des sorties graphiques et quantitatives. Enfin, une dernière étape exporte en format PDF et .xlsx l'ensemble des graphiques, cartes et tableaux produits. Le format PDF a été choisi pour sa dimension vectorielle et sa simplicité d'utilisation, ainsi que pour sa capacité à être facilement transformé en image non vectorielle par l'utilisateur. Le format .xlsx a été choisi de manière à permettre à l'utilisateur un retraitement des informations dans un tableur, et parce qu'il peut à la fois être aisément transformé en fichier .csv par l'utilisateur et lu nativement par un grand nombre d'outils de traitement de données. Cette étape d'export correspond en réalité à différents scripts R, avec un script par étape de production de sortie, de manière à permettre à l'utilisateur de ne pas exporter l'ensemble des visualisations mais uniquement celles qu'il veut conserver. Les graphiques et cartes sont exportés dans un répertoire Graphs et les tables dans un répertoire Tables; tous les exports sont dotés d'une nomination automatique, reposant sur un *prompt* soumis à l'utilisateur au début du script-maître (MASTER_R_form_distri_txt_SRC.R), tel qu'illustré en (3.15).

(3.15) *Prompt* de nomination soumis à l'utilisateur

```
1 | > Enter the name of the POS you want to investigate,
   |   to name the files:
2 | ADJ_femi_sg
```

Ici, si l'utilisateur répond ADJ_femi_sg (l. 2), le nom de tous les fichiers produits commencera par cette chaîne, ce qui permet un classement automatique de la dizaine de graphiques, cartes et tables produite. Une vue synthétique de l'ensemble des sorties qui peuvent être produites pour une catégorie morphosyntaxique est illustrée, pour les adjectifs féminins singuliers, en table 3.6 *infra*. Dans cette table, les sorties qui peuvent être produites individuellement ne sont pas indiquées (p.ex., ADJ_femi_sg_map_ALL.pdf, qui regroupe les cartes synchroniques de la distribution des taux de Ø devant #C et #V, donne aussi lieu à l'export individuel de chaque carte, produisant les fichiers ADJ_femi_sg_map_C.pdf et ADJ_femi_sg_map_V.pdf).

Du point de vue de l'utilisateur. À cette étape également, l'utilisation du programme demeure assez simple, ce qui réduit le risque d'erreur de manipulation. L'utilisateur ouvre le script-maître MASTER_R_form_distri_txt_SRC.R, importe l'ensemble des packages nécessaires et répond aux *prompts* qui lui sont soumis (3.16) et qui définissent le nom des fichiers, le contenu de certains textes imprimés sur les sorties graphiques et tabulaires ainsi que le comportement de certains scripts. Il exécute ensuite la série de scripts qui correspond au type d'analyse qu'il veut effectuer, ces différents types d'analyses étant documentés dans le script.

TABLE 3.6 – Ensemble des sorties graphiques, cartographiques et tabulaires de Comparalem. Sauf mention du contraire (tx #V; préval. élis.), toutes les sorties représentent les taux de Ø et/ou les taux de <e> finaux. Ici, « préval. élis. » abrège *prévalence de l'élision* et « synchro. » et « diachro. » abrègent respectivement [*représentation*] *synchronique* et *diachronique*.

Nom du fichier	Type	Contexte	Cf. fig.
Graphiques			
ADJ_femi_sg __C_lemma.pdf	diachro. par lemme	#C	5.3
ADJ_femi_sg __V_lemma.pdf	diachro. par lemme	#V	5.2
ADJ_femi_sg C_V.pdf	diachro.	#C vs #V	3.2
ADJ_femi_sg E_nonE.pdf	diachro. tx #V	-Ø vs <e>	3.3
Cartes			
ADJ_femi_sg map_ALL.pdf	synchro.	#C vs #V	3.6
ADJ_femi_sg map_elision_weight_t1-t2.pdf	diachro. préval. élis. (1100-1200)		3.10
ADJ_femi_sg map_elision_weight_t3-t4.pdf	diachro. préval. élis. (1200-1322)		3.11
ADJ_femi_sg map_elision_weight.pdf	synchro. préval. élis.		3.7
ADJ_femi_sg map_V_t1-t2.pdf	diachro. (1100-1200)	#V	3.8
ADJ_femi_sg map_V_t3-t4.pdf	diachro. (1200-1322)	#V	3.9
Tables			
ADJ_femi_sg_no_lemma.xlsx	synchro. et <e>	#C vs #V	3.5
ADJ_femi_sg.xlsx	synchro. par lemmes et par occ.		3.4

Synthèse du flux de travail de la seconde étape. La seconde étape d'utilisation de Comparalem consiste en l'exécution du script MASTER_R_form_distri_txt_SRC.R. Celui-ci, après avoir importé les données, applique plusieurs filtres : le premier vient de la sélection des lemmes effectuée à l'étape précédente, le deuxième filtre vient de l'exclusion non plus de lemmes mais d'occurrences repérées à l'étape précédente, le troisième ne sélectionne que les occurrences correspondant à des propriétés morphosyntaxiques annotées définies par l'utilisateur, et le quatrième retransche les textes situés après 1322.

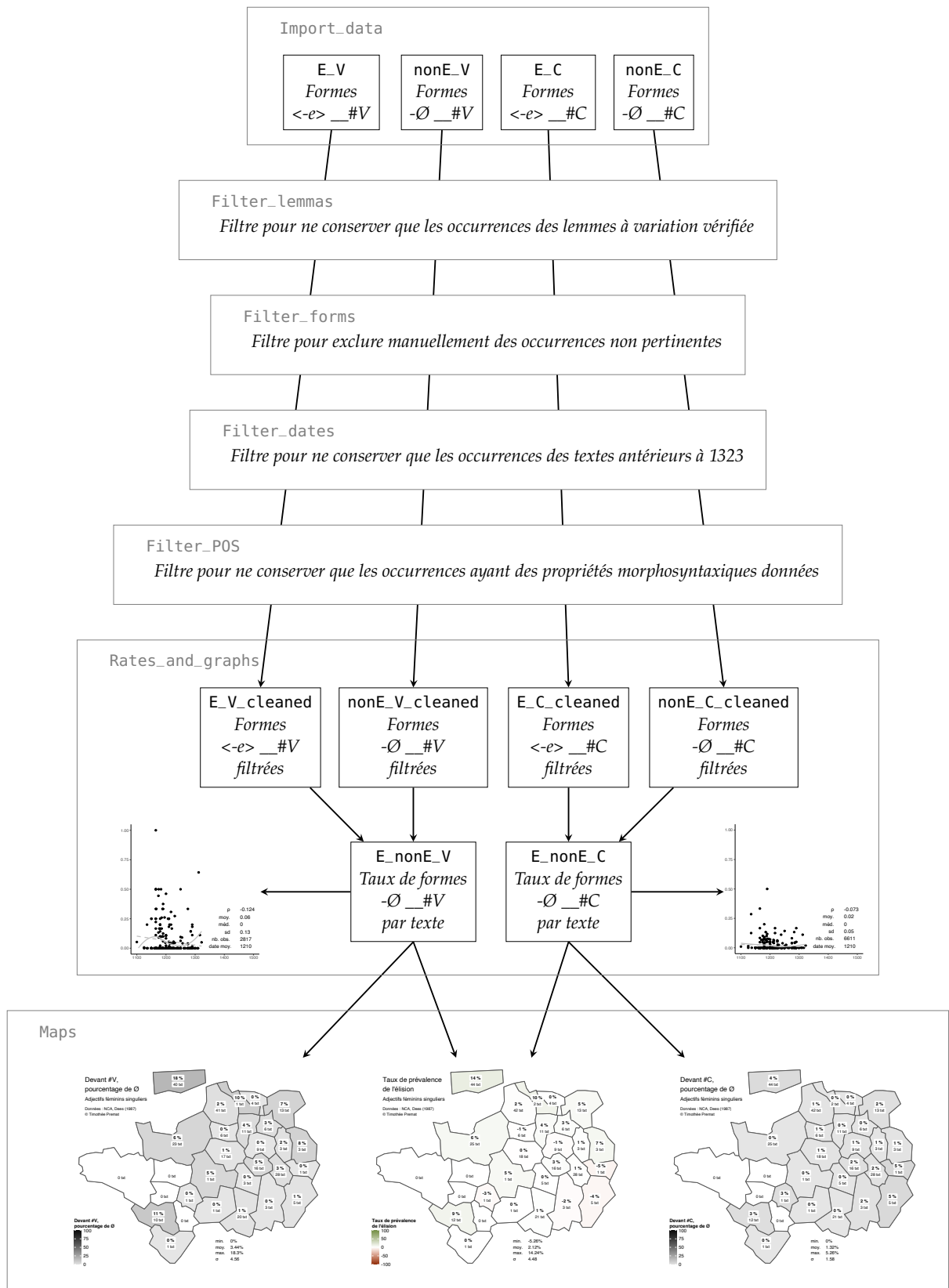
Ensuite, le programme effectue des opérations sur ces jeux de données pour calculer le taux de variantes -Ø pour chaque texte et propose des sorties d'abord graphiques et tabulaires, puis cartographiques, en synchronie puis en diachronie. Enfin, l'ensemble des sorties est exporté et enregistré sur la machine de l'utilisateur. Le fonctionnement d'ensemble de cette phase de travail de Comparalem est résumé en fig. 3.12, où les étapes de production des cartes diachroniques et d'export des sorties ne sont pas mentionnées.

(3.16) *Prompt* soumis à l'utilisateur

```

1 > Enter the name of the POS you want to investigate,
   to name the files:
2 > Enter the name of the POS you want to print as
   title:
3 > Enter the variable you want to use to filter forms
   . Can be gender_number, or gender_number_case:
4 > Will you by-pass filtering by lemmas? Answer Y or
   N:

```



Les cartes diachroniques ne sont pas représentées ici.

FIGURE 3.12 – Vue synthétique du flux de travail du deuxième temps d'exécution de Comparalem.

3.2.2 Note sur la reproductibilité et la diffusion du programme

Comme exposé *supra*, Comparalem n'est pas encore prêt à une diffusion publique, dans la mesure où son code et ses instructions sont encore spécialisés pour l'usage précis que nous en avons fait pour cette thèse. Néanmoins, à des fins de reproductibilité des analyses, le programme est disponible sur un dépôt GitHub.⁵ Le code est commenté et, avec les informations données dans ce chapitre et dans le chapitre précédent, un utilisateur familier du langage R devrait être capable d'exécuter le programme. Ce répertoire ne contient pas l'ensemble des données utilisées ni l'ensemble des sorties produites ; il contient, à titre d'exemple, les données issues de TXM pour les adjectifs, et les sorties produites par le traitement des adjectifs féminins singuliers. Nous projetons, dans le futur, le développement et la diffusion d'une version du programme qui soit optimisée et adaptée à l'étude de toute variation graphique.

5: https://github.com/TimotheePremat/Comparalem_alpha/.

3.2.3 En somme

Comparalem est un ensemble de scripts R qui permet d'étudier la distribution de deux variantes graphiques d'un même lemme, pour tous les lemmes. C'est là sa force principale : le programme exécute une comparaison systématique de tous les lemmes, pour observer dans un premier temps quels sont les lemmes qui sont affectés par la variation graphique en question, et, dans un second temps, comment sont distribuées les variantes dans l'espace et le temps. Il propose un certain nombre de sorties graphiques, cartographiques et tabulaires, ainsi que des mesures statistiques. Il prend comme *inputs* principaux des fichiers CSV issus de requêtes CQP de type Index effectuées dans TXM et propose un certain nombre de filtres sur ces données. À deux reprises, l'utilisateur a l'opportunité d'agir sur les données pour exclure des lemmes ou des occurrences et améliorer la spécificité du test conduit. Ce programme vise donc à nous permettre de dépasser la simple observation de formes variables, pour atteindre l'évaluation de leur représentativité et de leur distribution, et assurer ainsi à nos analyses linguistiques une assise empirique plus large.

3.3 Le PAM, Programme d'Analyse Métrique

Le PAM, *Programme d'Analyse Métrique* (POGGIO et PREMATA 2019 ; PREMATA et POGGIO [à paraître]), est un programme Python que nous avons développé avec Enzo Poggio, chercheur indépendant en informatique pour sciences humaines. Contrairement à Comparalem, le PAM est un programme d'annotation : il prend en *input* un texte versifié et pratique un découpage syllabique, une annotation des syllabes en fonction de leur position vis-à-vis de l'accent, et en déduit le nombre de syllabes métriques du vers et les possibles coupes que celui-ci peut comporter. Il a été conçu dans le but d'étudier les régularités et irrégularités de l'élosion des voyelles finales atones, par la détection de vers en apparence hypo- ou hyper-métriques. Le PAM possède des points communs avec le programme développé par Thomas Rainsford pour sa thèse de doctorat (RAINSFORD 2011b), mais

cette similarité tient essentiellement à une inspiration dans la démarche ; le code du PAM est tout à fait indépendant du code de Rainsford, et la philosophie, les heuristiques et le but des deux programmes diffèrent.

Avant de rentrer dans la description du fonctionnement du PAM, il importe de dire quelques mots à propos de son rôle dans ce travail. Ce programme a été conçu afin de détecter des irrégularités métriques qui montreraient une tendance diachronique, mais la phase exploratoire a révélé que, pour une écrasante majorité des textes, une telle tendance n’existait pas. En effet, dans la plupart des textes, la métrique est régulière et les irrégularités dans l’élision tiennent surtout aux cas où des monosyllabes fonctionnels, dont l’élision est régulière mais non systématique, ne sont pas élidés (comme *que* par exemple). Comme le PAM travaille sur le postulat que l’élision des monosyllabes est transcrite dans la graphie, il ne cherche pas à élider ceux-ci. Aussi, le principal apport du PAM, qui ne tient qu’en quelques lignes au chapitre 13, est de montrer que l’élision métrique est régulière. Néanmoins, dans le cadre d’une recherche annexe (PREMAT et POGGIO [à paraître]), le PAM nous a permis de travailler sur les élisions métriques non transcrites dans la graphie, dans la *Vie de saint Alexis* et dans *Roland*. Aussi, toute la dimension technique que couvre la présente section n’apparaît pas directement dans cette thèse, ayant simplement confirmé en négatif la régularité métrique d’une écrasante majorité de textes.

3.3.1 Flux de travail

Le PAM est un script Python qui s’exécute dans un terminal. Il est accompagné par des scripts compagnons : deux scripts, regroupés dans le projet `pam_corpus`, permettent de préparer les textes pour obtenir le format d’*input* du PAM, tandis qu’un troisième script, `pamviz.R`, permet la correction de certaines computations du PAM et l’obtention de sorties graphiques. Le PAM en lui-même procède principalement en trois étapes, dont la division est invisible pour l’utilisateur. Dans un premier temps, il produit un découpage syllabique du texte, avant de produire une annotation prosodique partielle de ce texte, vers par vers, puis une analyse métrique de ces propriétés. Il affiche ses résultats dans le terminal et peut également les exporter dans un ou plusieurs fichiers.

3.3.1.1 Préparation des textes : `pam_corpus`

En amont de l’utilisation du PAM, les données doivent prendre la forme de fichiers `.txt`, avec en principe un fichier par texte. Dans ce format, toutes les annotations doivent avoir été retirées, et les frontières de vers doivent être marquées par des retours chariot. Le projet `pam_corpus` permet de préparer automatiquement ces fichiers à partir du format XML dans lequel les corpus sont distribués. Il comporte deux scripts Python ; le premier est destiné au nettoyage des textes, le second à la gestion des chiffres romains. Ce sous-projet est distribué hors du PAM, et est accessible sur GitHub,⁶ il fonctionne pleinement sur les textes de la BFM et fonctionne en partie sur les textes du NCA, sans que nous ne l’ayons encore testé sur tous les textes de ce dernier corpus.

PREMAT et al. ([à paraître]). « Restituer des fragments de phonologie avec le PAM (*Programme d’Analyse Métrique*) : l’élision de schwa et des voyelles pleines dans les monosyllabes fonctionnels en très ancien français ».

AlexisRaM : ANON. (1050*). *Vie de saint Alexis*. Compo. ca 1050, ms. ca 1120, dial. norm. (RAINSFORD et MARCHELLO-NIZIA 2018).

roLand : ANON. (ca 1100). *Chanson de Roland*. Compo. ca 1100, ms. ca 1120, dial. norm. (MOIGNET 1972a).

6: https://github.com/EPgg92/pam_corpus.

Xml_txt.py. Le script `xml_txt.py` procède à l'élimination de toutes les balises d'annotation (y compris les préambules des fichiers), pour ne conserver que les formes, séparées par des espaces, et réunit les différents mots d'un même vers sur une seule ligne. Son format de sortie correspond donc aux fichiers que le PAM prend en *input*. Il s'exécute dans un terminal par la commande indiquée en (3.17), où `[input file]` et `[output file]` doivent être remplacés par le nom (et, le cas échéant, le chemin) des fichiers.

```
(3.17) python3 xml_txt.py [input file] > [output file]
```

Clear_numbers.py. Le script `clear_numbers.py` s'exécute après `xml_txt.py`, et a pour objectif de remplacer tous les chiffres romains par des chiffres écrits en toutes lettres. Nous avons adopté la graphie du français contemporain pour ces chiffres, pour ne pas adopter arbitrairement une graphie médiévale qui ne saurait correspondre à tous les textes. Le remplacement est automatique pour tous les cardinaux supérieurs à 1; pour 1, le programme comporte une interface qui permet à l'utilisateur de décider si le chiffre doit être fléchi au masculin (*un*) ou au féminin (*une*): il présente à l'utilisateur chaque occurrence de l'abréviation « i » et l'utilisateur indique, à partir du contexte, si cette occurrence est masculine ou féminine, en saisissant 1 ou 2 dans la ligne de commande. Cette étape est nécessaire, dans la mesure où le féminin est bisyllabique tandis que le masculin ne l'est pas. Le programme s'exécute dans un terminal par la commande indiquée en (3.18); il ne prend pas d'indication de fichier de sortie, et place automatiquement ses sorties dans un répertoire intitulé `clear_txt`. L'exécution du script déclenche la mobilisation de l'interface pour la sélection de la flexion de *un(e)*, telle qu'illustrée en (3.19).

```
(3.18) python3 clear_numbers.py [input file]
```

```
(3.19) Interface de flexion de un(e)
```

```
1 | corpus/BFM/YvainKu.txt line: 154
2 | Einz me leissasse .i. des danz traire ,
3 | Choose 1:un or 2:une
```

3.3.1.2 Exécution du PAM

Une fois les corpus préparés, l'utilisateur peut passer à l'exécution du PAM proprement dit. Dans la commande d'exécution du programme, l'utilisateur désigne le texte, active s'il le souhaite des options d'export et d'affichage, fournit au PAM ce que nous appelons le nombre de syllabes métriques de référence, dénoté $m[n]$ (p.ex. $m[8]$ pour un texte en octosyllabes, $m[10]$ pour un texte en décasyllabes). La première opération que réalise le programme est de procéder à la syllabation du texte, c'est-à-dire au découpage des syllabes. Cette étape est représentée par un script autonome, `syllabise.py`. Une fois la syllabation effectuée, le PAM peut attribuer une étiquette relative à l'accent pour chaque syllabe et, à partir de la syllabation, de cet étiquetage et du nombre de syllabes métriques de référence, il peut en déduire les propriétés métriques du vers, et ce à deux niveaux : au niveau du nombre et du statut des syllabes, et au niveau de possibilités de coupes à l'intérieur du vers. À la fin de l'exécution, le PAM

affiche ses résultats dans le terminal et, si l'utilisateur a demandé un ou plusieurs exports dans sa commande d'exécution, le PAM enregistre le ou les exports demandés.

Syllabise.py. La première étape est donc celle de la syllabation. Elle est assurée par un script *ad hoc*, qui est également disponible de manière autonome sur GitHub.⁷ Ce script repose notamment sur la librairie NLTK (BIRD, KLEIN et LOPER 2009). Le syllabeur repose sur une base graphique, et opère selon des paramètres que nous avons établis à partir de notre connaissance de la syllabation de l'ancien français et avons affinés progressivement au vu des résultats. Il dispose, dans un dictionnaire au format `yaml`, de la liste des graphèmes vocaliques et des graphèmes consonantiques, ainsi qu'une liste d'exceptions courantes : groupes consonantiques tautosyllabiques, diphtongues, etc. Pour chaque voyelle ou diphtongue, le syllabeur génère une syllabe, et distribue les consonnes autour de la voyelle, en préférant les attaques aux codas. Les frontières de syllabes internes aux mots sont marquées, dans le texte, par des points médians « · ». Il s'agit donc d'une approche dotée d'un *a priori* : le programme dispose d'instructions de syllabation que nous avons définies. D'autres approches, plus poussées, permettent d'éviter un tel *a priori* (voir p.ex. RAINSFORD et SCRIVNER (2014)), mais les résultats de `syllabise.py` sont tout à fait satisfaisants pour notre approche. L'effet du script est illustré en (3.20) sur le premier vers de *roland*, la version de la BFM de la *Chanson de Roland*.

(3.20) car·les li reis nost·re em·pe·re·re ma·gnes

On peut ici observer que les diphtongues sont bien prises en charge (reis et non re·is), que la préférence pour les attaques plutôt que pour les codas fonctionne (em·pe·re·re), tout en excluant les attaques qui n'appartiennent pas à la liste des groupes tautosyllabiques ; en revanche, le programme a échoué à répartir correctement les consonnes de *nostre* (nost·re au lieu de nos·tre). Ce type d'imprécision est ici toléré dans la mesure où le placement des consonnes en coda interne n'est, dans le PAM, qu'une question d'affichage et ne change rien à l'annotation prosodique et à l'analyse métrique. Enfin, il faut noter que le script produit une annotation bornée par les frontières de mots, ce qui signifie qu'il est insensible aux élisions qui ne seraient pas transcrites dans la forme graphique des mots ; c'est dans le PAM, *via* l'annotation, que cette dimension de la concaténation des mots sera prise en charge.

Annotation prosodique. Une fois le texte syllabé, le PAM procède à une annotation de chaque syllabe, en fonction du caractère accentuable,⁸ prétonique ou posttonique de la syllabe. Il attribue donc l'une des étiquettes présentées en (3.21) à chaque syllabe, en distinguant entre d'une part les syllabes posttoniques sans coda et devant initiale vocalique, et d'autre part les syllabes posttoniques dotées d'une coda et/ou situées devant une initiale consonantique.

7: <https://github.com/EPgg92/syllabise.py>.

BIRD et al. (2009). *Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit*.

RAINSFORD et al. (2014). « Metrical annotation for a verse treebank ».

roLand: ANON. (ca 1100). *Chanson de Roland*. Compo. ca 1100, ms. ca 1120, dial. norm. (MOIGNET 1972a).

8: *Syllabes accentuables* et *syllabes accentuées* diffèrent en ce que l'une est au statut de potentialité tandis que l'autre est au statut d'actualisation. Une fois l'accent de groupe advenu, l'accent devient assigné à un domaine qui dépasse le mot lexical : on peut déduire de la forme d'un mot et de sa catégorie grammaticale s'il dispose d'une syllabe accentuable et, si c'est le cas, quelle est cette syllabe, mais c'est la position du mot au sein du domaine qui transforme ces syllabes accentuables en syllabes accentuées ou non.

(3.21) Étiquettes d'annotation prosodique du PAM.

- 1. Est étiquetée -1 toute syllabe posttonique sans coda et située devant initiale vocalique ;
- 0. Est étiquetée 0 toute syllabe posttonique disposant d'une coda et/ou située devant un mot à initiale consonantique, ainsi que toute syllabe appartenant à un clitique ;
- 1. Est étiquetée 1 toute syllabe prétonique ;
- 2. Est étiquetée 2 toute syllabe accentuable.

La détection du statut des syllabes repose sur un fondement exclusivement graphique : lorsque le mot est monosyllabique, le PAM étiquette sa syllabe 2 (3.22 a). Lorsque le mot est polysyllabique, le PAM définit si la dernière syllabe a pour noyau un <e> et ne dispose pas de coda (sauf <s> et <nt>) (→ 0 ou -1 : 3.22 b) ou si cette dernière syllabe a pour noyau un autre graphème vocalique ou un <e> suivi d'une autre coda que <s> ou <nt> (→ 2 : 3.22 c). Si la dernière syllabe a pour noyau un <e>, le programme cherche si ce <e> est suivi de <s> ou <nt> et/ou d'une consonne quelconque à l'initiale du mot suivant ; si c'est le cas, la syllabe est annotée 0 et le <e> est pourvu d'un tréma, si ce n'est pas le cas, la syllabe est étiquetée -1 (3.22 d). Il annote ensuite 2 toute syllabe précédant une syllabe étiquetée -1 ou 0 (3.22 e), et annoté 1 toute syllabe précédant une syllabe étiquetée 2 (3.22 f).

(3.22) Exemple d'association des étiquettes

- (a) 2. re²is ; ple²ins ; mè²r ;
- (b) {0, -1}. emper^{0, -1}ère ; no^{0, -1}strè ; car^{0, -1}lès ;
- (c) 2. est²èt ; con²quist ;
- (d) 0 vs -1. emperer⁰è (*magnes*) ; no⁻¹strè (*emperere*) ; carl⁰ès
- (e) 2. emper²èrè ; no²strè ; car⁻¹lès ;
- (f) 1. emper¹èrè ; est¹èt ; con²quist.

Les étiquettes présentées en (3.21), et dont l'association aux syllabes est illustrée en (3.22), correspondent *a priori* à des statuts métriques, tels que détaillés en (3.23). C'est cette correspondance qui sous-tend l'analyse métrique de seconde étape. L'effet de cette annotation peut être observé sur les trois premiers vers de roland en (3.13). Comme illustré dans cet exemple, l'annotation prosodique du PAM se place *au-dessus* du texte, chaque coupure de mot dans le texte étant reprise dans l'annotation par une barre verticale. Le numéro du vers est également affiché à droite du texte. Comme énoncé *supra*, les syllabes élidées ne sont démarquées des autres syllabes que par leur étiquette -1, leur syllabation issue de `syllabize.py` ne différant pas de celle des autres syllabes.

(3.23) Étiquettes d'annotation prosodique du PAM.

- 1. Syllabe élidée
- 0. Syllabe métrique interne ou initiale, ou extramétrique finale⁹
- 1. Syllabe métrique interne
- 2. Syllabe métrique accentuable (peut être la proéminence de l'hémistiche ou du vers).

9: Il y a là une évolution du programme depuis la publication de POGGIO et PREMAT (2019) ; dans la version 2019 du PAM, les rimes féminines étaient également annotées -1, mais ce n'est plus le cas depuis la version 2020.


```

1 | 2 0 | 0 | 2 | 1 -1 | 1 1 2 0 | 2 0
2 | car·les li reis nost·re em·pe·re·rē ma·gnes num_l:1
3 | 2 | 2 | 0 | 2 | 2 | 1 2 | 0 | 1 2 0
4 | set anz tuz pleins ad es·tet en es·pai·gne num_l:2
5 | 0 | 0 | 0 | 2 | 1 2 | 0 | 2 -1 | 1 2 0
6 | tresqu en la mer cun·quist la te·re al·tai·gne num_l:3

```

FIGURE 3.13 – *Chanson de Roland*, v. 1-3, analyse du PAM.

Par ailleurs, le PAM a accès à deux dictionnaires de formes au format `yaml`, l'un contenant des syllabations (`special_syll.yaml`), et l'autre contenant des annotations prosodiques (`special_type.yaml`). Si la forme que le PAM traite appartient au dictionnaire de syllabation, le PAM ignore la syllabation issue de `syllabise.py` et recopie celle indiquée dans le dictionnaire; si la forme appartient au dictionnaire d'annotation, le PAM s'abstient de pratiquer son analyse et recopie simplement l'annotation indiquée dans le dictionnaire. Ces dictionnaires sont conçus dans une double visée : d'une part, ils contiennent déjà un grand nombre de formes courantes qui échappent aux principes de syllabation et/ou d'annotation par défaut, et, d'autre part, ils sont conçus pour être aisément édités par l'utilisateur. Ainsi, c'est dans le dictionnaire d'annotation que sont stockés l'ensemble des adverbes en *-ment* : comme ceux-ci se terminent par un <e> suivi de <nt>, le PAM les analyse de la même manière qu'il analyse les terminaisons verbales de 3pl, et les annote comme étant des paroxytons (...2 0). Nous avons donc extrait dans TXM la liste des adverbes en *-ment* et avons forcé leur annotation oxytone (...1 2) dans `special_type.yaml`. C'est également dans ce dictionnaire d'annotation que figurent les clitiques, de manière à ce que ceux-ci soient annotés 0 et non 2. Ainsi, en (3.13), les monosyllabes *li*, *tuz*, *en*, *tresqu* et *la* sont annotés 0 puisqu'ils sont présents dans le dictionnaire d'annotation `special_type.yaml`, qui spécifie pour eux l'étiquette 0; leur étiquette par défaut aurait été 2, comme pour tous les monosyllabes (cf. *reis*, *anz*, etc.).

Le dictionnaire d'étiquetage a été nourri à partir de requêtes CQP dans TXM (nous avons ainsi extrait l'ensemble des déterminants, des prépositions, etc.) et de manière manuelle à partir d'erreurs constatées. Les détails des choix qui ont été opérés au niveau de l'inclusion dans ce dictionnaire de certaines formes peuvent être discutés; ces choix sont principalement mus par une volonté pragmatique quant à l'obtention d'un proxy de la structure métrique qui soit à même de fonctionner pour le plus grand nombre de vers possible. Il n'est pas toujours possible de décider, sur la base d'une simple forme graphique (et donc sans accès à la position du mot au sein du domaine), si le mot est atone interne (0) ou tonique final (2). Ainsi, nous n'avons pas annoté 0 les cardinaux, puisque, s'ils apparaissent en déterminants comme dans *set anz tuz pleins* en (3.13), ils peuvent aussi apparaître en fonction adjectivale avec omission du nom auquel il se rapportent, et peuvent alors porter l'accent; ces deux positions sont illustrées pour *set milie* (« sept mille »)¹⁰ par les deux vers en (3.24). Dans la mesure où l'annotation 0 0 de *milie* aurait provoqué la détection d'un vers hypométrique lorsque *milie* termine un vers (3.24 b)

10: La graphie *milie* correspond, dans la *Chanson de Roland*, non pas à « millier » mais à « mille » : dans la dernière syllabe, le <i> note soit la palatalisation de /l/ en /ʎ/ soit un /j/ précédant le schwa (SHORR 2007, p. 103). Le traitement métrique du cardinal dans le texte est univoque : *milie* est toujours traité comme un paroxyton bisyllabique dépourvu de coda, dont la posttonique peut être élidée, ou être extramétrique en coupe épique ou à la rime, cf. (3.24 b).

(cf. *infra*). Cela permet de travailler facilement sur tous les vers dont le nombre de syllabes métrifiées n'est pas conforme au mètre de référence $m[]$, et donc de repérer facilement les irrégularités dans le vers et, dans notre cas, celles qui sont dues à un traitement irrégulier d'une voyelle finale atone.

Analyse métrique de second niveau. La mécanique décrite dans la sous-section précédente ignore les coupes. Elle est suffisante pour traiter les vers simples (non sécables en hémistiches), mais pas les vers complexes. Le PAM dispose donc, en sus des mécanismes décrits ci-dessus, d'un mécanisme de repérage et d'annotation hors-ligne¹¹ des coupes. Le programme évalue toutes les frontières de syllabes situées entre la troisième position métrique (i.e. la troisième syllabe sans compter les syllabes annotées -1) et l'antépénultième position métrique. Cette évaluation prend deux critères : (a) la présence ou l'absence d'une frontière de mot, et (b) l'étiquette prosodique de la syllabe qui précède cette coupe.

Cette évaluation retourne l'une des 5 valeurs données en (3.25), augmentée d'un chiffre qui correspond à la position métrique de la syllabe précédent la coupe (p.ex. 4ma signifie : coupe masculine après la quatrième position métrique). Nous référons ici à la position donnée par ce chiffre par le mot *locus*, et la signalons par une barre oblique '/' située au terme de la syllabe qui correspond à cette position.¹² *Locus-1* et *locus+1* renvoient aux syllabes situées respectivement avant et après la syllabe précédant la coupe. La flèche à corps vide '⇒' signifie conventionnellement un alignement, ici dans l'étiquetage prosodique : '*locus* ⇒ 2' signifie que la syllabe située avant la coupe est étiquetée comme étant tonique (valeur 2, cf. 3.21 *supra*). Une glose de l'algorithme est indiquée en marge.

(3.25) Conditions de l'étiquetage des coupes par le PAM.

ma Coupe masculine. Statut attribué si :

→ Frontière de mot : *locus*]_w,

→ *Locus* ⇒ 2,

ex. : *Carles li reïs/ nostr²ē empererē magnes* 4ma (v. 1)

épC Coupe épique devant consonne. Statut attribué si :

→ Frontière de mot : *locus+1*]_w ,

→ *Locus* ⇒ 2,

→ *Locus+1* ⇒ 0,

ex. : *Li reis marsi/ liē la tient ki deu nen aimet* 4épC (v. 7)¹³

épV Coupe épique devant voyelle. Statut attribué si :

→ Frontière de mot : *locus+1*]_w,

→ *Locus* ⇒ 2,

→ *Locus+1* ⇒ -1,

ex. : *Li reis marsi/ liē esteit en sarraguce* 4épV (v. 10)

ly Coupe lyrique. Statut attribué si :

→ Frontière de mot : *locus*]_w,

→ *Locus* ⇒ 0,

→ *Locus-1* ⇒ 2,

ex. : *Il nē sevēnt/ liquels els la veintrat* 4ly (v. 735)

11: Pour des raisons de lisibilité et de distinction entre l'annotation prosodique et l'analyse métrique, l'annotation des coupes ne s'effectue pas au sein de l'annotation prosodique, mais en péricoré. Il s'agit, en effet, de dissocier strictement ce qui tient de l'annotation des unités *présentes dans le texte* et ce qui tient de l'analyse de la disposition métrique de ces unités, cette dernière étant par essence *absente du texte*.

12: Nous utilisons ici des notations mixtes : nous ne reproduisons l'annotation du PAM qu'autour de la coupe et n'indiquons les hiatus entre une tonique et sa posttonique qu'en ce lieu. Nous notons l'apocope régulière du schwa final devant initiale vocalique par 'ϕ' et son maintien régulier devant initiale consonnantique par 'è'.

ma. Pour qu'une coupe masculine soit valide, il faut qu'un oxyton soit situé immédiatement avant la position de coupe.

épC. Pour qu'une coupe épique devant initiale consonnantique soit valide, il faut qu'un paroxyton soit situé à la coupe, celle-ci tombant entre la tonique et la posttonique du paroxyton, et que le mot suivant commence par une consonne, ce qui interdit l'élision de la posttonique (étiquette 0).

13: *Marsilie* est un paroxyton de trois syllabes, cf. note 10 *supra*.

épV. Pour qu'une coupe épique devant initiale vocalique soit valide, il faut qu'un paroxyton soit situé à la coupe, celle-ci tombant entre la tonique et la posttonique du paroxyton, et que le mot suivant commence par une voyelle; le PAM détecte une élision (étiquette -1).

ly. Pour qu'une coupe lyrique soit valide, il faut qu'un paroxyton soit situé immédiatement avant la coupe, et que la voyelle posttonique soit suivie par une consonne, ce qui interdit son élision (étiquette 0).

ej Coupe enjambante (ou à l'italienne). Statut attribué si :

→ Frontière de mot : $locus+1]_w$,

→ $Locus \Rightarrow 2$,

→ $Locus+1 \Rightarrow \emptyset$,

ex. : *Li emperè/res² tent ses mains vers deu* 4ly (v. 137)

NA 'Non-attribuable'. Statut attribué si aucune des étiquettes précédentes n'a pu être attribuée à ce *locus*.

ex. : *Repairet s'en¹/a joiϕ e a barnage* 4NA (v. 3944)

ej. Pour qu'une coupe enjambante soit valide, il faut qu'un paroxyton soit situé à la coupe, celle-ci tombant entre la tonique et la posttonique du paroxyton, et que la voyelle posttonique soit suivie par une consonne, ce qui interdit l'éliision de la posttonique (étiquette \emptyset).

Les exemples donnés en (3.25) pour les trois dernières coupes le sont uniquement dans le but d'illustrer le fonctionnement de l'algorithme. La *Chanson de Roland* n'admettant ni coupe lyrique ni coupe enjambante, l'analyse qui doit être retenue n'est pas celle produite en (3.25); les vers 735 et 136 doivent plus probablement être analysés comme comportant une coupe masculine après la sixième position métrique. De la même manière, le vers 3944 doit être analysé comme comportant une coupe épique devant initiale consonantique après la sixième position métrique. Comme le PAM scanne tous les *loci* situés entre la troisième et l'antépénultième positions métriques, il propose plusieurs possibilités de coupes, comme illustré en (3.26), où c'est bien la seconde coupe proposée par le PAM qui doit être retenue.

(3.26) Multiplicité des coupes

v. 735	<i>Il nē se²vent/liquē²ls/ els la veintrat</i>	4ly/6ma
v. 137	<i>Li emperè/res² tent/ ses mains vers deu</i>	4ej/6ma
v. 3944	<i>Repairet s'en¹/a joi·ϕ²/ e a barnage</i>	4NA/6épV

Il faut ici noter que le PAM ne fait que proposer des coupes, sans jamais en sélectionner une. De ce fait, le programme ne prend pas en compte la possibilité d'une coupe épique devant initiale consonantique dans le calcul du nombre *m* de syllabes métriques, ce qui fait que tous les vers dotés d'une telle coupe sont hypermétriques, comme l'est le v. 7 en (3.25 épC), dont l'affichage, reproduit en (3.27), indique un vers anisosyllabique (*m*: 11[10]). De ce fait, les vers dotés d'une coupe épique devant initiale consonantique épC sont toujours détectés comme hypermétriques par le PAM.

(3.27) Vers faussement hypermétrique avec coupe épC : roland, v. 7

1	\emptyset 2 1 2 \emptyset \emptyset 2 \emptyset 2 2 2 \emptyset	<i>m</i> : 11[10]	3NA/4épC/5NA/6NA/7NA
2	li reis mar·si·liē la tient ki deu nen ai·met	<i>num_l</i> : 7	

De même, les posttoniques en coupe épique devant initiale vocalique épV, si elles ne changent pas le nombre de syllabes métriques *m*, demeurent étiquetées -1 comme si elles étaient élidées, alors qu'elles sont plus probablement prononcées mais extramétriques, comme dans le v. 10 en (3.25 épV), dont l'affichage complet est reproduit en (3.28). Dans cet exemple, la finale atone *-lie* (dont le noyau est un schwa, cf. note 10 *supra*) de *Marsilie* est annotée comme élidée (-1), dans la mesure où le PAM n'a pas accès à l'analyse des coupes lorsqu'il pratique l'annotation

des syllabes. Or, si les coupes épiques devant initiale consonantique indiquent que la posttonique à la coupe est bien prononcée mais est extramétrique, il n’y a aucune raison qu’il en aille autrement pour les coupes épiques devant initiale vocalique (cf. sec. 13.2.2.2).

(3.28) Syllabe faussement élidée (-1) en coupe épV : roland, v. 10

1	0	2		1	2	-1		1	2		0		1	1	2	0	m:10[10]	3NA/4épV/5NA/6ma/7NA
2	li reis mar·si·lie es·teit en sar·ra·gu·ce															num_l:10		

Affichage des résultats. Par défaut, le PAM affiche dans le terminal où il est exécuté l’annotation et le texte de chaque vers, suivis par une vue d’ensemble des propriétés métriques du texte. Cet affichage, appelé *sommaire*, est reproduit en (3.29) pour roland. Il présente le nombre et la proportion des vers détectés comme bien formés ($m = m[]$) et mal-formés ($m \neq m[]$), en général (l. 2-3), et pour chaque valeur de m détectée (l. 4-11). Dans la deuxième partie du sommaire, sont affichés le nombre et la proportion des vers détectés comme disposant potentiellement d’une césure, pour chaque type de césure et chaque position (l. 13-43).

L’utilisateur peut, lors de la commande d’exécution du programme, spécifier une commande d’affichage et des commandes de sélection de vers, telles que détaillées en (3.30). Hormis la première commande, ces instructions permettent de ne travailler que sur un sous-ensemble de vers au sein du texte, en les sélectionnant par numéro de vers (3.30 b-d), par nombre de syllabes métriques détectées (3.30 e-f) et/ou par type de coupes proposées (3.30 g-h). Cela permet notamment de contrôler les résultats du PAM. C’est par exemple en vérifiant tous les vers hyper- ou hypo-métriques que nous avons ajusté les dictionnaires `special_syll.yaml` et `special_type.yaml`.

Ces commandes sont cumulatives : si l’utilisateur indique `-g -c 4épC -t 11`, il ne disposera que de l’affichage du sommaire (-g) concernant les vers qui contiennent une potentielle coupe épique devant initiale consonantique après la quatrième position métrique (-c 4épC) et qui ont aussi été détectés comme comportant un 11 syllabes métriques (-t 11). Dans le cas de roland, les commandes `-c 4épC -t 11` et `-c 4épC` génèrent le même résultat, puisque tous les vers disposant d’une potentielle coupe épique devant une initiale consonantique après la quatrième position métrique ont été détectés par le PAM comme faisant 11 syllabes métriques. Le résultat de ces commandes est reproduit en (3.31). En revanche, la commande `-t 11` ne produit pas le même résultat, puisque 205 vers annotés comme contenant 11 syllabes métriques ne comportent pas de possibilité d’appliquer une coupe épique après la quatrième position métrique (3.32). Ainsi, la lecture de (3.32) permet de considérer que 80% des 1026 vers que le PAM a détecté comme faisant 11 syllabes métriques n’en font, en réalité, que 10, puisqu’ils sont dotés d’une coupe épique.

(3.29) Affichage du sommaire dans le PAM (roland)

```

1 1 ../../pam_corpus/clear_txt/roland.txt m:10
2 70.28% soit 2809/3997 vers bien formés
3 29.72% soit 1188/3997 vers mal formés
4   - m:7[10]  00.05% soit  2
5   - m:8[10]  00.10% soit  4
6   - m:9[10]  01.28% soit 51
7   - m:11[10] 25.67% soit 1026
8   - m:12[10] 02.28% soit  91
9   - m:13[10] 00.25% soit  10
10  - m:14[10] 00.08% soit  3
11  - m:15[10] 00.03% soit  1

13 Distribution des césures pour 3997 vers:
14   - 3NA  85.09% soit 3401
15   - 3ej  01.13% soit  45
16   - 3ly  03.18% soit 127
17   - 3ma  09.73% soit 389
18   - 3épC 00.13% soit  5
19   - 3épV 00.75% soit  30
20   - 4NA 11.83% soit 473
21   - 4ej  01.00% soit  40
22   - 4ly  01.13% soit  45
23   - 4ma 60.92% soit 2435
24   - 4épC 20.54% soit 821
25   - 4épV 04.58% soit 183
26   - 5NA 80.86% soit 3232
27   - 5ej  02.45% soit  98
28   - 5ly  01.00% soit  40
29   - 5ma 15.49% soit 619
30   - 5épC 00.10% soit  4
31   - 5épV 00.10% soit  4
32   - 6NA 63.07% soit 2521
33   - 6ej  07.73% soit 309
34   - 6ly  02.45% soit  98
35   - 6ma 24.12% soit 964
36   - 6épC 01.50% soit  60
37   - 6épV 01.13% soit  45
38   - 7NA 57.22% soit 2287
39   - 7ej  07.91% soit 316
40   - 7ly  07.73% soit 309
41   - 7ma 22.92% soit 916
42   - 7épC 02.93% soit 117
43   - 7épV 01.30% soit  52

```

(3.30) Commandes d'affichage et de sélection du PAM

- (a) -g Affiche uniquement le sommaire ;
- (b) -n Sélectionne uniquement les vers dont le numéro de ligne correspond à l'argument fourni après -n ;
- (c) -a Sélectionne uniquement les vers dont le numéro de ligne est supérieur ou égal à l'argument fourni après -a ;
- (d) -b Sélectionne uniquement les vers dont le numéro de ligne est inférieur à l'argument fourni après -b ;
- (e) -t Sélectionne uniquement les vers dont le nombre de syllabes métriques *m* correspond à l'argument fourni après -t ;
- (f) -T Sélectionne uniquement les vers dont le nombre de syllabes métriques *m* ne correspond pas à l'argument fourni après -T ;
- (g) -c Sélectionne uniquement les vers dont l'une des coupes correspond à l'argument fourni après -c ;
- (h) -k Sélectionne uniquement les vers dont aucune des coupes ne correspond à l'argument fourni après -k.

(3.31) Restriction de l'analyse aux vers *m*:11 et 4épC (roland)

```

1 | 1 ../../pam_corpus/clear_txt/BFM/roland.txt m:10
2 | 00.00% soit 0 /821 vers bien formés
3 | 100.00% soit 821 /821 vers mal formés
4 | - m:11[10] 100.00% soit 821
6 | Distribution des césures pour 821 vers:
7 | - 3NA 100.00% soit 821
8 | - 4épC 100.00% soit 821
9 | - 5NA 100.00% soit 821
10 | - 6NA 96.59% soit 793
11 | - 6épC 03.41% soit 28
12 | - 7NA 89.16% soit 732
13 | - 7épC 10.84% soit 89

```

(3.32) Restriction de l'analyse aux vers *m*:11 (roland)

```

1 | 1 ../../pam_corpus/clear_txt/BFM/roland.txt m:10
2 | 00.00% soit 0 /1026 vers bien formés
3 | 100.00% soit 1026/1026 vers mal formés
4 | - m:11[10] 100.00% soit 1026
6 | Distribution des césures pour 1026 vers:
7 | - 3NA 99.51% soit 1021
8 | - 3épC 00.49% soit 5
9 | - 4NA 19.98% soit 205
10 | - 4épC 80.02% soit 821
11 | - 5NA 99.61% soit 1022
12 | - 5épC 00.39% soit 4
13 | - 6NA 94.15% soit 966
14 | - 6épC 05.85% soit 60
15 | - 7NA 88.60% soit 909
16 | - 7épC 11.40% soit 117

```

Exports. L'utilisateur peut, lors de l'exécution du programme, spécifier un ou plusieurs exports. Ces exports, toujours bornés par les éventuelles restrictions du type (3.30 b - h), sont de deux types : d'une part, le programme peut sauvegarder l'équivalent de ce qu'il a affiché dans le terminal, c'est-à-dire l'annotation et le texte de chaque vers puis le sommaire, ou uniquement le sommaire si l'option `-g` a été utilisée. Cet export se fait en format tabulaire (`.csv` ou `.xlsx`) ou en format texte (`.txt`). D'autre part, le programme peut ne sauvegarder que les informations métriques, dans une table au format `.txt` contenant une seule ligne par vers et disposant de colonnes pour le numéro du vers, son nombre `m` de syllabes métriques détectées, et ses possibilités de césure. Alors que le premier export permet de lire le texte et son annotation, ce n'est pas le cas du second export, qui a en revanche l'avantage d'offrir une seule ligne par vers, alors que le premier type d'export affiche, comme dans le terminal, deux lignes par vers : l'annotation se trouve sur une ligne au-dessus du texte syllabé. Cette présentation en deux lignes, pratique pour la lecture humaine, est en revanche peu adaptée à un traitement informatique des données issues du PAM. Ces exports sont déclenchés, par l'ajout à la commande d'exécution du programme de `-S` suivi du format (`csv`, `xlsx` et/ou `txt`) pour l'export sur deux lignes, et de `-L` pour l'export sur une ligne. Dans le cas de la reproduction de l'affichage du terminal, le fichier prend le nom du texte suivi de la date et du moment de l'exécution de la commande (p.ex. `roland.2023-01-01 12/30/59.xlsx`); dans le cas de l'export des seules propriétés métriques, le fichier prend le nom du texte suivi de l'indication `line_by_line_meter` (p.ex. `roland_line_by_line_meter.txt`). Ces fichiers sont placés dans un répertoire `pam_output` au sein du répertoire `script` contenant le cœur du programme.

Du point de vue de l'utilisateur. Une fois les fichiers préparés pour leur utilisation dans le PAM, l'utilisateur doit simplement disposer d'une fenêtre de terminal pointant sur le répertoire `script` où se trouve le PAM, à l'intérieur du répertoire `pam` dans le dépôt disponible sur GitHub. Il lui suffit ensuite de rédiger une commande commençant par la commande d'exécution du programme (`./pam.py`), suivie de l'indication du fichier cible précédé de `-f`, du nombre `m[n]` de syllabes de références, précédé de `-m`, suivi d'éventuelles options d'affichage, de restriction de l'analyse et d'export. Ces commandes sont illustrées en (3.33) pour un usage simple et en (3.34) avec un affichage du seul sommaire, avec une restriction aux vers détectés comme ayant 11 syllabes métriques et une possible coupe épique après la quatrième position métrique, et avec deux options d'export.

```
(3.33) ./pam.py -f
        ../../pam_corpus/clear_txt/BFM/roland.txt
        -m 10
```

```
(3.34) ./pam.py -f
        ../../pam_corpus/clear_txt/BFM/roland.txt
        -m 10 -g -c 4epC -t 11 -S csv -L
```


3.3.2 Configurations linguistiques spécifiques : les *config*.

Comme indiqué *supra*, le PAM repose sur deux dictionnaires qui permettent de corriger la syllabation et l'annotation de formes données. Ces dictionnaires, situés dans le répertoire *config*,¹⁴ représentent une configuration linguistique du PAM : en plus des principes contenus dans le script *syllabise.py* et dans les scripts du PAM lui-même, ces dictionnaires permettent d'ajuster facilement le rapport du programme à certaines formes linguistiques. À ces dictionnaires s'ajoutent, dans le répertoire *config*, le dictionnaire général réglant le rapport entre graphèmes et phonèmes (liste des voyelles, consonnes, diphtongues, groupes tautosyllabiques), dénommé *constants.yaml*. L'utilisateur a non seulement la possibilité d'éditer les trois dictionnaires par défaut, mais il peut aussi créer de nouveaux dictionnaires, pour adapter le programme aux spécificités linguistiques d'un texte donné. Dans ce cas, l'utilisateur crée un nouveau répertoire contenant ces trois fichiers, les édite, et indique dans la commande d'exécution du PAM qu'il veut recourir à une autre configuration linguistique que celle par défaut, au moyen de l'instruction `-C` suivie du chemin menant vers le répertoire contenant les trois nouveaux dictionnaires.

14: `pam/script/config` dans l'architecture de dossiers telle que disponible sur le dépôt GitHub.

De fait, la *config*. par défaut correspond principalement à la langue de Chrétien de Troyes, à partir de laquelle nous avons établi les dictionnaires principaux. Dès qu'un texte s'éloigne de cette langue, il peut être nécessaire d'adapter la *config*.. Ainsi, par exemple, la *Vie de saint Alexis* (`alexisRaM`) contient quelques occurrences de voyelles finales atones notée <a> là où l'ancien français n'a que <e> : *estra* (v. 95) ~ *estre* ; *tuta la cuntretha* (v. 20) ~ *tute la contree*, etc. Dans certains cas, ces formes paroxytones peuvent être homographes de formes oxytones dans d'autres textes, ainsi *prendra* (INF dans *il la volt prendra*) se confond avec l'INDFUT3sg *prendra*. Il faut donc éditer, non pas les fichiers généraux de *config*., mais uniquement des fichiers spécifiques à ce texte. En l'occurrence, l'édition de ces exemples s'opère par l'ajout de lignes sur le modèle de celles en (3.35). Ici, les étiquettes à appliquer à chaque syllabe sont données chronologiquement sous la forme d'une liste en-dessous de la forme du mot.

(3.35) Édition de `special_type.yaml`

```

1 | prendra:
2 |   - 2
3 |   - 0
4 |
5 | estra:
6 |   - 2
7 |   - 0
8 |
9 | cuntretha:
10 |   - 1
11 |   - 2
12 |   - 0
```

Il en va de même pour l'ensemble des consonnes finales archaïques qui disparaissent en ancien français mais sont encore présentes dans *Roland* et *Alexis*, comme <t> dans *abandunet* (v. 390, INDPRÉ3sg), *adubet*

(v. 1987, *id.*). Ces formes doivent être annotées comme des paroxytons (leur <et> final représente /ət/), alors que la configuration par défaut du PAM les analyse comme des oxytons, ne tolérant que <s> et <nt> après un schwa final. Pour couvrir ces formes, nous avons procédé à une recherche, dans TXM, de toutes les formes verbales se terminant par <et>, les avons vérifiées une par une et avons ensuite corrigé l’annotation des formes paroxytoniques, sur le modèle de (3.35), de manière à obtenir une annotation se terminant par 2 0.

L’utilisateur peut donc, assez facilement, établir une configuration linguistique qui corresponde au texte ou à l’ensemble de textes sur lequel il travaille, en modifiant un ou plusieurs des dictionnaires et en faisant appel à cette nouvelle *config.* dans la commande d’exécution du PAM. Par exemple, pour analyser le texte roland en fonction de la *config.* que nous avons préparée pour le très ancien français, la commande serait telle que reproduite en (3.36).

```
(3.36) ./pam.py -f
        ../../pam_corpus/clear_txt/BFM/roland.txt
        -m 10 -C ../config/config_tufr
```

3.3.3 Post-traitement des résultats en R

Le PAM est accompagné par un script R, `pamviz.R`, qui permet d’une part d’appliquer des opérations sur les résultats du PAM, et d’autre part d’obtenir une vue d’ensemble des données de chaque texte analysé, par la production de sorties graphiques. Si nous utilisons dans la partie métrique de ce traitement certains des résultats de ce script, d’autres ne sont utilisés que pour contrôler l’analyse et ne sont pas présents dans le texte de cette thèse. Ce script, dont le développement n’est pas finalisé mais qui est opérationnel, est disponible sur un dépôt GitHub autonome.¹⁵

Les sorties de ce script présentent deux avantages principaux sur les données de sortie du PAM : d’une part, le script permet la correction de la détection du nombre *m* de syllabes métriques des vers dotés d’une coupe épique devant initiale consonantique (qui passent, par exemple, de *m*:11[10] à *m*:10[10] pour *Roland*). D’autre part, ce script propose un ensemble de sorties qui permettent de contrôler différents aspects de la distribution des longueurs de vers détectées. Ce contrôle s’effectue dans quatre sorties différentes, dont la dernière regroupe plusieurs visualisations.

15: <https://github.com/TimotheePremat/PhD/blob/main/MetricalAnalysis/pamviz.r>.

3.3.3.1 Rectification de *m*

La première opération que permet ce script est la rectification du nombre *m* de syllabes métriques des vers qui sont analysés comme disposant d’une potentielle coupe épique : dans le cas d’un texte en décasyllabes (*m*[10]), tous les vers détectés comme disposant d’une coupe épique après la quatrième ou la sixième position métrique (4épC ou 6épC) et dont le nombre de syllabes métriques est égal à 11 (*m*:11[10]) voient leur longueur *m* corrigée à 10 (3.37), puisque la coupe épique implique l’extramétricalité de la syllabe posttonique à la coupe.

$$(3.37) \quad \left[\begin{array}{c} m: 11[10] \\ \left\{ \begin{array}{c} 4\text{épC} \\ 6\text{épC} \end{array} \right\} \end{array} \right] \rightarrow \left[\begin{array}{c} m: 10[10] \\ \left\{ \begin{array}{c} 4\text{épC} \\ 6\text{épC} \end{array} \right\} \end{array} \right]$$

L'effet de l'opération formalisée en (3.37) est implémenté au début du script, et ses résultats sont donc reflétés dans l'ensemble des productions de ce script. L'utilisateur obtient ainsi, pour les textes autorisant la coupe épique, un affichage équivalent à la première partie du sommaire du PAM (3.29), mais avec un nombre de vers de longueur canonique et de vers hypermétriques par une syllabe qui intègre le traitement de l'extramétrie des syllabes posttoniques en coupe épique. Ainsi, en table 3.7, le nombre de vers dont la longueur est canonique ($m: 10[10]$) atteint 3662 (92%), alors qu'il n'atteignait que 2809 (70%) dans la sortie du PAM qui ignorait l'extramétrie des posttoniques en coupe épique (3.29). Le nombre de vers hypermétriques par une syllabe ($m: 11[10]$) est diminué d'autant : il est de 173 vers (4%) après traitement de l'extramétrie des coupes épiques, contre 1026 (26%) en sortie du PAM.

TABLE 3.7 – Sortie tabulaire de `pamviz.R` pour *Roland*. Table issue de PREMAY et POGGIO ([à paraître]).

meter	count	rate
7	2	0.05
8	4	0.10
9	51	1.28
10	3662	91.62
11	173	4.33
12	91	2.28
13	10	0.25
14	3	0.08
15	1	0.03

3.3.3.2 Distribution des nombres de syllabes métriques

La première sortie graphique propose, sous la forme d'un histogramme, de contrôler la distribution générale des longueurs des vers m du texte, c'est-à-dire des nombres de syllabes métriques par vers détectées par le PAM, après la correction (3.37) des hypermétries déclenchées par l'absence de sensibilité de m à la présence d'une coupe épique. Sur ce graphique, l'échelle de l'axe des ordonnées est logarithmique, de manière à rendre visibles les différences entre les longueurs de vers non canoniques, qui seraient écrasées par la masse des vers de longueur canonique dans une échelle régulière. Ce type de graphiques est illustré en fig. 3.16 pour *Roland*. Sur cette figure, on observe une distribution régulièrement décroissante du nombre de vers à mesure que leur longueur m s'éloigne de la longueur de référence $m[10]$. Si la distribution ne suit pas exactement une loi normale, cet effet est en partie attendu, étant donné qu'on ne saurait considérer qu'un vers hypermétrique et un vers hypométrique sont strictement équivalents sur le plan de la genèse de l'anisosyllabisme, et que l'hypométrie semble toujours moins marquée que l'hypermétrie, ne pouvant évoluer vers zéro de la même manière que l'hypermétrie évolue par augmentation du nombre m de syllabes métriques.¹⁶

16: Différentes approches statistiques de cette asymétrie peuvent être envisagées. Pour l'utilisation d'un coefficient de symétrie γ_1 et l'hypothèse d'une symétrie par défaut (implication de la loi normale) des distributions de longueurs de vers, voir PREMAY et POGGIO ([à paraître]).

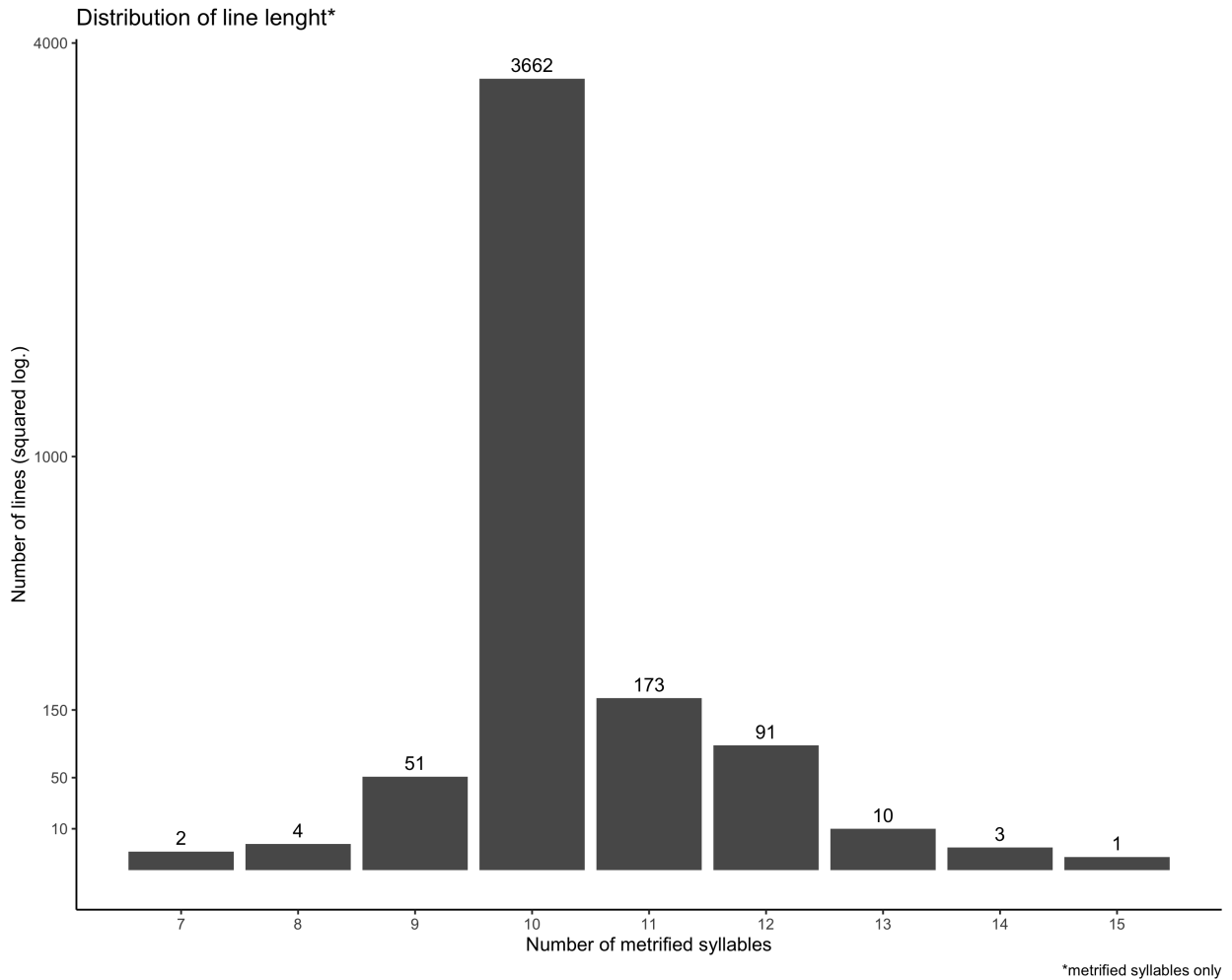


FIGURE 3.16 – Distribution des longueurs de vers m de *Roland*. Analyse produite par le PAM, avec adaptation de la *config.* au très ancien français. Figure issue de PREMAT et POGGIO (à paraître).

3.3.3.3 Position des vers dans le texte

La seconde sortie graphique permet de contrôler la position des vers en fonction de leur longueur m dans le texte. Elle prend la forme d'un nuage de points, où chaque point représente un vers, et où le déroulement du texte est représenté en abscisse, tandis que l'ordonnée représente le nombre de syllabes métriques, après la correction (3.37) des hypermétres déclenchées par l'absence de sensibilité de m à la présence d'une coupe épique. Ce type de visualisation permet essentiellement de vérifier si certaines sections du texte sont plus sensibles que d'autres à l'anisyllabisme. Cela ne semble pas être le cas pour *Roland*, dont le graphique de position des vers est reproduit en fig. 3.17.

3.3.3.4 Relation entre schwas et longueur du vers

La dernière sortie graphique proposée par le script `pamviz.R` compile 8 graphiques qui interrogent la présence de schwas éliés et non éliés au

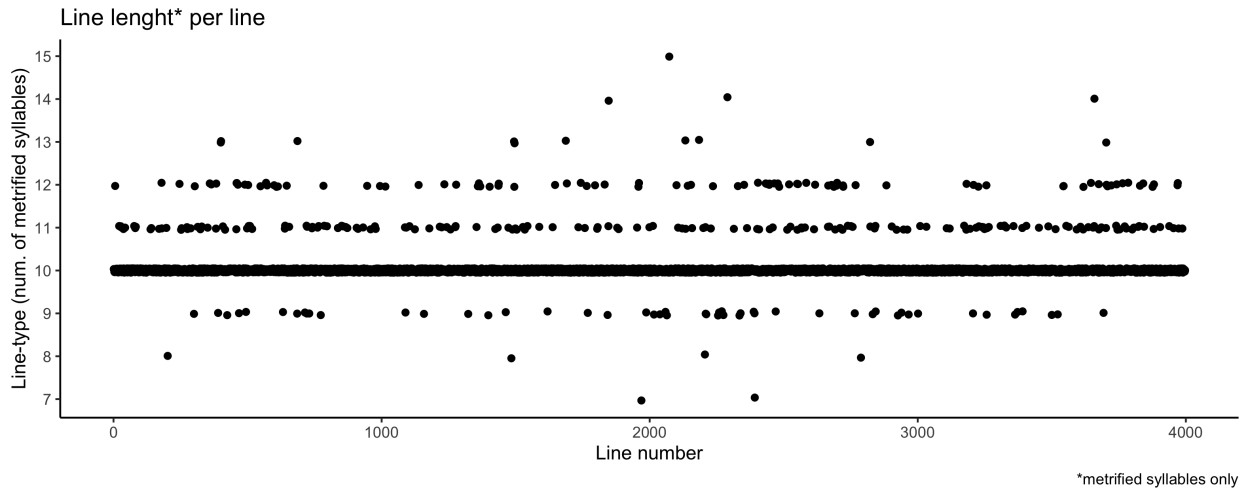


FIGURE 3.17 – Distribution des longueurs de vers m de *Roland* en fonction de la position des vers dans le texte. Analyse produite par le PAM, avec adaptation de la *config.* au très ancien français. Les points sont ‘secoués’ (en. *jittering*) vis-à-vis de leur position réelle pour représenter les superpositions.

sein des vers. Cette compilation est reproduite en fig. 3.18, et détaillée dans les paragraphes suivants. L'utilisateur peut évidemment extraire les graphiques indépendamment, que ce soit dans l'interface de son environnement R ou en ajoutant quelques lignes au script.

A. Quantité de schwas par vers. Les deux premiers graphiques, A.a et A.b, représentent simplement le nombre de schwas finaux contenus dans un vers, en fonction du nombre de vers. Dans le cas de *Roland*, la comparaison de A.a et A.b permet de visualiser le fait que si plus de la moitié des vers contiennent au moins un schwa non éli­dé (placé devant initiale consonantique ou extramétrique), seule une faible proportion (12%) de vers contiennent des schwas éli­dés (placés devant initiale vocalique). Cette mesure, si elle dépend essentiellement du contexte environnant, n'est pas orthogonale à notre analyse, dans la mesure où elle vient nuancer la quantité absolue des schwas éli­dés devant initiale vocalique dans la langue de *Roland*.

B. Quantité de schwas en fonction de la longueur des vers. Les deux graphiques B.a et B.b représentent également le nombre de schwas éli­dés par le PAM (placés devant initiale vocalique) ou non éli­dés (placés devant initiale consonantique) par vers, en fonction du nombre de vers, mais également en fonction de la longueur m des vers concernés. Elle permet de mettre en lumière que si le nombre de schwas non éli­dés est globalement semblable pour les vers hypermétriques par une syllabe ($m: 11[10]$) et pour les vers isosyllabiques ($m: 10[10]$), ce n'est pas le cas pour les vers hypermétriques par plus d'une syllabe ($m>11$) (B.a). Si, évidemment, la probabilité de présence d'un schwa augmente avec l'augmentation de la longueur du vers, le fait que ce ne soit pas le cas des schwas non éli­dés pour les vers hypermétriques par une syllabe ($m: 11[10]$) et que ce ne soit pas le cas pour les schwas éli­dés est un indice intéressant.

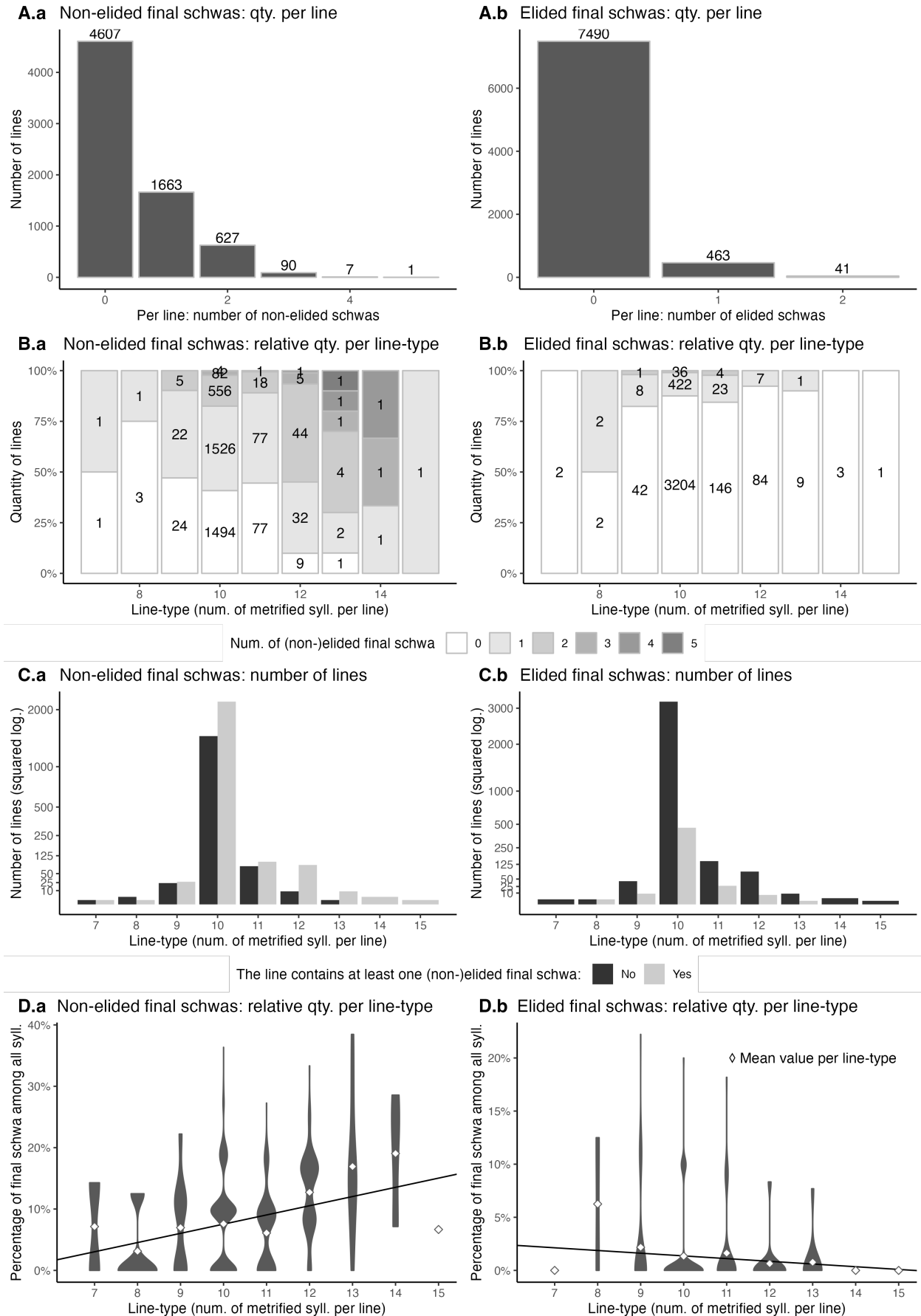


FIGURE 3.18 – Représentation de différents aspects de la relation entre schwas, longueurs des vers et nombre des vers dans *Roland*. Analyse produite par le PAM, avec adaptation de la *config.* au très ancien français.

C. Présence de schwas en fonction du nombre de vers. Les deux graphiques C.a et C.b représentent la présence ou absence d'au moins un schwa dans un vers, en fonction de la longueur du vers et du nombre de vers concernés. Ils reprennent donc les informations des graphiques B.a et B.b, en résumant le nombre de schwas par vers à un critère binaire 'présence/absence' de schwa(s) dans le vers, mais en représentant graphiquement le nombre de vers concernés par chaque catégorie, ce qui n'était visible sur les graphiques B.a et B.b que sous la forme d'une étiquette numérique. Comme sur le graphique général de distribution des longueurs de vers (fig. 3.16), l'échelle des ordonnées est logarithmique. Ces graphiques permettent de contrôler visuellement les mêmes données que les graphiques B.a et B.b, mais avec cette fois le poids absolu des longueurs de vers représentées. La différence de nombre de schwas non élidés entre les vers hypermétrique par une (m: 11 [10]) et deux syllabes (m: 12 [10]) est confirmée en C.a, tandis que l'absence de cet effet pour les schwas élidés visible en B.b est reproduite en C.b.

D. Vue d'ensemble. Les graphiques D.a et D.b représentent, avec des graphiques à forme de violon (en. *violin plots*), le taux de schwas non élidés (a) et élidés (b) parmi les syllabes des vers, en fonction de la longueur des vers (ordonnées) et en fonction du nombre de vers concernés (largeur du 'violon'). L'effet est intéressant, l'hypothèse nulle prédisant un taux constant de schwas peu importe la longueur du vers. Dans *Roland*, il apparaît que plus un vers est hypermétrique, plus son taux de schwas non élidés est élevé, en revanche, pour les schwas élidés, l'effet apparaît nul : non seulement la pente est réduite, mais elle est principalement due aux deux vers de huit syllabes métriques contenant deux schwas élidés.

3.3.3.5 En somme

Le script `pamviz.R` qui accompagne le PAM permet la rectification du compte des vers hypermétriques à coupe épique, et propose un certain nombre de visualisations graphiques qui permettent de contrôler la distribution des différentes longueurs de vers, en quantité et en position dans le texte, ainsi que les liens potentiels entre la quantité de schwas présents dans un vers et l'anisosyllabisme de celui-ci. Nous ne ferons pas grand usage de ces graphiques dans ce travail, mais ils ont été systématiquement consultés avant l'analyse des résultats du PAM, d'une part pour contrôler des biais qui seraient inhérents au texte ou à son traitement, et d'autre part pour contrôler certains effets systématiques. Ces visualisations ne seront donc mobilisées que ponctuellement, là où elles apporteront à l'analyse un argument intéressant.

3.3.4 Note sur la reproductibilité et la diffusion du programme

Le programme PAM, les scripts de préparation du corpus et le script `pamviz.R` sont accessibles sur des dépôts GitHub. Ils sont néanmoins à des niveaux de développement différents, aussi le niveau de reproductibilité qu'ils assurent varie.

Le PAM. Le développement du PAM a atteint une étape de stabilisation, mais la diffusion des *config*. reste à améliorer au sein du dépôt GitHub. Il est commenté et propose, lorsqu'il est exécuté dans un terminal avec l'ajout d'un argument `-h`, un résumé des différentes options possibles, pour guider l'utilisateur dans la rédaction de la commande d'exécution. De fait, et avec l'aide des ressources ici développées et présentes dans deux publications (POGGIO et PREMAT 2019; PREMAT et POGGIO [à paraître]), un utilisateur habitué à un usage minimal de la ligne de commande devrait parvenir à utiliser le PAM de façon satisfaisante. Nous considérons donc que le PAM satisfait aux conditions de reproductibilité des analyses ici déployées, même si la diffusion des *config*. reste à améliorer.

<https://github.com/EPgg92/pam>.

—**Syllabize.py.** Le syllabeur `syllabize.py`, qui est utilisé par le PAM, est disponible dans une version stable mais imparfaite, sur un dépôt Github autonome. Le principal problème de cette version du syllabeur est que celui-ci ne prend pas en compte les élisions, dans la mesure où celles-ci sont traitées dans l'annotation du PAM (-1) et non dans la syllabation de la forme graphique des mots. Ce problème, inhérent au fait que le syllabeur effectue la syllabation d'une forme graphique (et non d'une forme phonologique ou phonétique), ne peut pas être résolu sans repenser l'architecture générale du programme. Nous avons néanmoins choisi de rendre le syllabeur disponible. La version du syllabeur intégrée au PAM ne dispose pas de cette limite, puisque le PAM intègre l'élision dans l'annotation; les conditions de reproductibilité sont ici pleinement remplies.

<https://github.com/EPgg92/syllabise.py>.

Pam_corpus. Le projet `pam_corpus`, qui contient les deux scripts python de préparation du corpus, est stabilisé et son développement est considéré comme achevé. D'utilisation simple, il est disponible en ligne et correspond aux exigences de reproductibilité pour préparer les textes à analyser avec le PAM. Il est pleinement fonctionnel sur l'ensemble des textes de la BFM, et fonctionne au moins sur une partie des textes du NCA (nous n'avons pas pu le tester sur tous les textes de ce dernier corpus).

https://github.com/EPgg92/pam_corpus.

pamviz.R. Le script `pamviz.R` est encore en cours de développement, et n'est diffusé qu'à titre provisoire. Il se contente, au demeurant, d'appliquer quelques opérations sur les données et de produire des visualisations. Il est disponible sur un dépôt Github. Le script, tel que livré sur GitHub, cible les données issues de l'analyse de `roland`. Les conditions de reproductibilité sont remplies, à condition que l'utilisateur modifie le programme pour cibler d'autres fichiers, ce que l'on peut attendre de la part d'un utilisateur à même d'exécuter un script R.

<https://github.com/TimotheePremat/PhD/blob/main/MetricalAnalysis/pamviz.r>.

3.3.5 En somme

Le PAM est un programme qui vise au repérage des vers anisosyllabiques. Il est accompagné par trois scripts : les deux premiers, appartenant au projet `pam_corpus`, permettent la préparation des textes en amont de l'exécution du PAM, tandis que le troisième script, `pamviz.R`, permet une exploitation statistique des résultats du PAM.

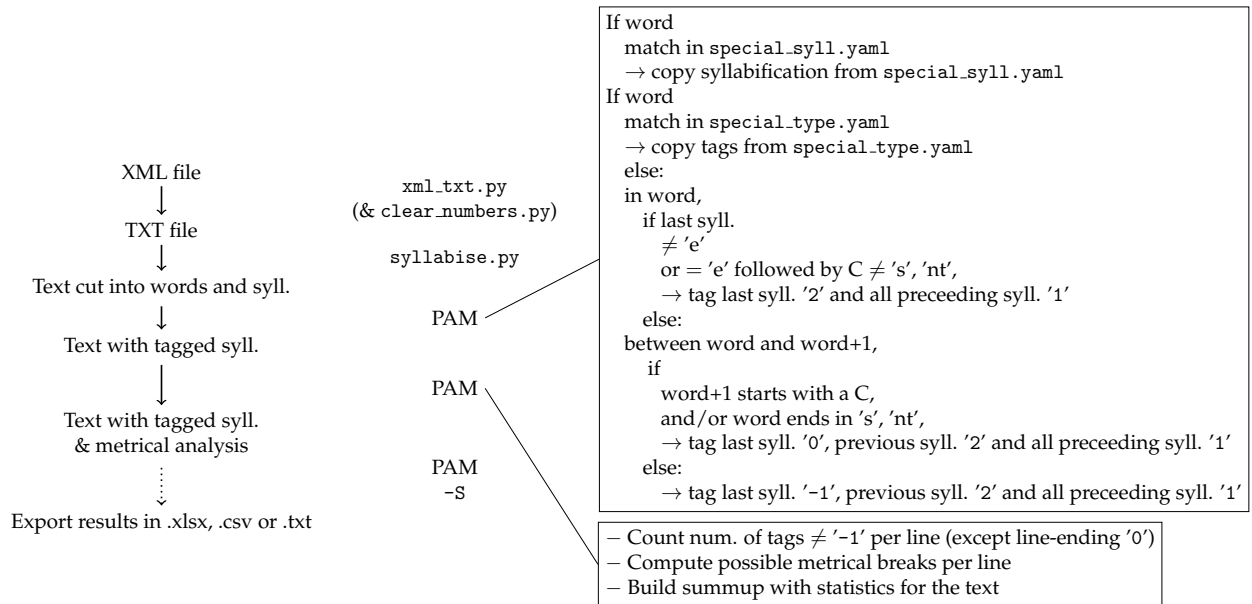


FIGURE 3.19 – Vue synthétique du flux de travail du PAM, comprenant la préparation des textes par les scripts du projet `pam_corpus`, mais ne comprenant pas le script de sorties graphiques `pamviz.R`. Figure issue de PREMAY et POGGIO (à paraître).

Une fois les textes nettoyés de leurs balises par les scripts du projet `pam_corpus`, le PAM procède par étapes : il utilise le script `syllabize.py` pour découper le texte en syllabes, puis annote ces syllabes en fonction de leur statut accentuel et propose une analyse métrique que l'on peut décomposer en deux niveaux. Le premier niveau consiste à compter les syllabes métriques des vers, c'est-à-dire toutes les syllabes à l'exception des posttoniques qui sont élidées ou qui terminent le vers. Le second niveau consiste à calculer, par un simple algorithme logique (tel que détaillé en 3.25), les possibilités de coupe du vers. Enfin, le PAM affiche un résumé des différents types de vers présents dans le texte, ainsi que des possibilités de coupe. Les résultats de ces annotations et computations sont affichés dans le terminal où le programme est exécuté, et peuvent être exportés par l'utilisateur. Ce flux de travail est synthétisé en fig. 3.19.

Pour traiter l'élosion des schwas finaux sans coda devant initiale vocalique ainsi que pour calculer le nombre de syllabes métriques et les possibilités de coupe des vers, le PAM produit une annotation de chaque syllabe, en fonction du caractère posttonique, prétonique ou accentuable de la syllabe. Néanmoins, le PAM opérant sur un fondement purement graphique, cette annotation est imparfaite et le programme est incapable de gérer des homographes qui n'auraient pas le même statut accentuel. Dans la mesure où le but de cette phase d'annotation n'est pas de fournir une annotation prosodique exhaustive et précise mais de détecter les syllabes finales élidées et non élidées, de compter le nombre de syllabes métriques par vers et de calculer les possibilités de coupe en fonction de l'accent, cette faiblesse du programme ne représente pas un biais à même d'invalider les résultats du PAM quant au calcul des syllabes métriques. Cette limite peut néanmoins fausser le calcul des coupes, aussi, lorsque les possibilités de coupe font partie de notre analyse, celles-ci ont été vérifiées manuellement et exhaustivement.

Si le PAM propose un décompte des différentes longueurs de vers et des coupes, il ne propose pas de traitement statistique avancé de ces données, et ne propose pas de sorties graphiques. Cette étape, qui permet de dépasser le repérage des vers pour monter en généralité, est effectuée à l'aide du script `pamviz.R`, qui propose des sorties graphiques représentant non seulement les différentes longueurs de vers du texte, mais également certains aspects de l'interaction entre schwas et longueur des vers. Ce script permet aussi de corriger, selon le principe formalisé en (3.37), la longueur des vers dotés d'une coupe épique devant initiale consonantique.

Ainsi, le PAM, les outils dédiés à la préparation du corpus (`pam_corpus`) et au traitement statistique des résultats du PAM (`pamviz.R`) permettent un traitement systématique des vers d'un texte. Ils permettent non seulement de repérer tous les vers anisosyllabiques, mais également d'interroger la relation entre schwa et longueur des vers, pour en déduire s'il y a lieu de considérer que des défauts d'isosyllabismes indiquent une irrégularité dans le traitement du schwa final.

3.4 Conclusion

Dans cette thèse, nous ne visons pas à sélectionner quelques exemples isolés et à bâtir une analyse linguistique sur le présupposé de leur représentativité; nous voulons au contraire disposer de données quantifiables que l'on peut interroger en fonction d'un certain nombre de variables. Pour parvenir à ces fins, nous recourons à une approche de corpus, qui vise à quantifier, au sein de corpus définis, la distribution des différentes configurations recherchées. Cette démarche contient également la possibilité de la sérendipité : la démarche systématique sur corpus fait apparaître des régularités ou des irrégularités auxquelles le chercheur n'avait pas nécessairement pensé. Elle permet surtout d'interroger la représentativité des occurrences, ainsi que leur diffusion dans le temps et l'espace. Mais cette démarche, qui repose ici sur un traitement informatique, n'est possible que lorsque des corpus numériques satisfaisants en termes quantitatifs et qualitatifs existent; ainsi, nous n'avons pas pu l'adapter à l'approche de la musication, dont l'analyse reste essentiellement artisanale.

Cette thèse repose donc sur deux corpus principaux, le NCA et la BFM. L'utilisation de chaque corpus répond, dans ce travail, à un objectif précis : le NCA est utilisé pour l'étude des alternances graphiques avec zéro en position finale atone, tandis que la BFM est utilisée pour l'analyse métrique. Chaque type d'analyse est produit par un programme informatique développé dans le cadre de cette thèse : `Comparalem` vise à analyser la régularité des variations graphiques des syllabes posttoniques, tandis que le PAM vise à permettre l'analyse des régularités et irrégularités dans le traitement métrique de ces posttoniques.

Dans cette partie, nous avons détaillé en chapitre 2 la composition de chaque corpus utilisé, les informations contenues dans chaque corpus et les limites inhérentes à ces corpus. En chapitre 3, nous avons détaillé la conception, l'utilisation et les limites des programmes informatiques utilisés. Hormis à l'ouverture du prochain chapitre, qui propose une

illustration pratique de l'emploi de Comparalem, il ne sera plus question de cette dimension technique dans la suite de ce travail, l'essentiel de ce qui devait être dit à ce sujet ayant été couvert ici. Ainsi, la partie suivante, dédiée aux résultats de l'analyse graphique du NCA avec Comparalem, se focalise sur les résultats de l'analyse. Les résultats de l'analyse métrique avec le PAM et de l'exploration artisanale des corpus de musication seront, eux, mobilisés dans la partie de discussion de ces résultats, pour venir enrichir et nuancer les informations fournies par l'analyse graphique du NCA. En effet, si ces approches dépendent d'un niveau d'abstraction supplémentaire, celui de la paraphonologie, métrique et musicale, elles permettent néanmoins de dépasser le domaine de l'analyse graphique, qui ne peut, par définition, représenter qu'une partie des phénomènes phonologiques.

II

DÉPOUILLEMENT DES ALTERNANCES GRAPHIQUES

CETTE partie se focalise sur le dépouillement des alternances graphiques vocaliques en position finale atone, c'est-à-dire principalement les alternances entre <e> pour /ə/ et zéro, et les alternances entre d'autres graphèmes vocaliques et zéro pour certains monosyllabes fonctionnels. Pour le premier type, nous parlons de 'défaut de <e>' lorsque le graphème <e> devrait être présent en position posttonique pour une forme, mais qu'il fait défaut; de cette manière, un *défaut de <e>* se distingue d'un Ø final, dans la mesure où, alors qu'un défaut de <e> est toujours aberrant au regard de l'étymologie et de la phonétique historique, un Ø peut être de droit (p.ex. AURUM → *or*; PLANUS → a.fr. *plain*), et s'opposer à un <e> qui serait, lui, épenthétique (*ore*) ou issu d'un autre étymon (lat. pop. †*plánia* → *plaine*; cf. sec. 4.1.2). Cette analyse des défauts de <e> constitue le cœur empirique de ce travail, et s'effectue à l'aide du programme Comparalem abordé dans la partie précédente.

Cette partie est organisée selon les parties du discours interrogées, des catégories variant le moins aux catégories disposant de la plus forte variation. Nous traitons ainsi des défauts des noms (chap. 4), des adjectifs féminins singuliers (chap. 5), des prépositions (chap. 6), des adverbes (7), des articles (8), des pronoms personnels (chap. 9) puis des conjonctions (chap. 10). Cette partie se termine par un chapitre portant sur les réductions de hiatus à l'intérieur du mot et sur les syncopes, qui fournit un point de comparaison avec la situation des voyelles finales atones (chap. 11).

Dans chaque chapitre, nous procédons à une description des données concernant les défauts de graphème vocalique final, variable par variable, avant de discuter ces résultats pour passer des observations aux généralisations. L'ensemble des généralisations proposées dans ces chapitres est reproduit en annexes (p. 587) pour permettre une lecture d'ensemble de celles-ci, si nécessaire.

4.1 Introduction

Ce chapitre se concentre sur la variation <e>/∅ des noms. Dans ce travail, le mot *nom* est pris au sens restreint de *substantif*, à l'exclusion des adjectifs, donc. Les noms propres ne sont pas systématiquement exclus de cette étude, mais Comparalem ne permet la détection que d'un seul nom propre atteint d'une variation <e>/∅ finale qui ne tienne pas à l'alternance de patrons oxytons et paroxytons dans la déclinaison.

Comme les noms ne connaissent, en ancien français, qu'une seule voyelle posttonique, /ə/, l'alternance avec zéro ne concerne pas d'autres voyelles finales. Les quelques occurrences où /ə/ est noté par une autre voyelle, toujours <a>, sont écartées; elles sont considérées comme phonologiquement régulières (le rapport entre graphèmes et phonèmes de ces textes permet la notation d'un schwa final par un <a>), correspondent toujours au type en <e>, et sont donc sans impact sensible sur les résultats statistiques.

Dans la suite de cette introduction, après un avertissement sur la pauvreté des résultats de ce chapitre, nous reprenons la préparation des données pour Comparalem en l'illustrant par ce cas pratique (sec. 4.1.1) et, comme nous le faisons dans les chapitres suivants, traitons des différents types d'exclusions que nous avons appliquées à nos données (sec. 4.1.2). Nous passons ensuite à la présentation des résultats, qui démontre l'extrême marginalité de la variation <e>/∅ que Comparalem parvient à détecter (sec. 4.2), avant de proposer quelques remarques typologiques (4.2.1) et d'étudier les distributions diachroniques et diatopiques de ces défauts (4.2.2). Une brève section de généralisations conclut que les quelques occurrences détectées par le programme ne sont pas significatives (4.3).

Avertissement au lecteur

Les données de ce chapitre sont essentiellement inconclusives, ou du moins conclusives en négatif : nous ne trouvons aucun défaut de <e> significatif dans la classe des noms. S'il est toujours impossible de démontrer l'inexistence d'un objet — c'est le cygne noir de Juvénal,^a nous entendons néanmoins permettre au lecteur de suivre notre raisonnement pour qu'il puisse juger de l'absence de découverte.

a. JUVÉNAL, *Satires*, VI : *rara avis in terris nigroque simillima cygne*, « oiseau rare en ce pays, comme un cygne noir », en parlant d'une femme de qualité. Les cygnes noirs sont originaires de l'hémisphère Sud.

4.1.1 Rappel de méthode : Comparalem

Ce chapitre, comme les autres chapitres de cette partie, repose sur l'utilisation de Comparalem. Comme il s'agit ici du premier cas d'utilisation de Comparalem, nous reprenons rapidement la rédaction des requêtes puis

le traitement des données par Comparalem, pour donner au lecteur une meilleure compréhension de ce que représentent les valeurs utilisées.

4.1.1.1 Requêtes CQP

Nous avons utilisé Comparalem avec des requêtes CQP permettant, dans un premier temps, de distinguer les formes dotées d'un schwa final sans coda consonantique des formes sans schwa final, et, dans un second temps, de distinguer ces formes selon que le mot suivant est à initiale consonantique ou vocalique. Ce second type de requêtes est reproduit en (4.1-4.3).

- (4.1) NOM_E_V :
`[word=".*e" & pos="NOM.*" & lemma!=".*é" & lemma!=".*_é.*"] [word="(a|e|o|y).*"]`
- (4.2) NOM_nonE1_V :
`[word!=".*e(s|z)?" & pos="NOM.*"] [word="(a|e|o|y).*"]`
- (4.3) NOM_nonE2_V :
`[word=".*e" & pos="NOM.*" & lemma="[a-zäëïöüÿáéíóúàèìòùâêîôûç]*é"]`
`[word="(a|e|o|y).*"]`

La requête en (4.1) nous permet d'obtenir toutes les occurrences se terminant par un <e>, annotées comme étant un nom, dont le lemme (ou les lemmes dans le cas où Tree Tagger propose des lemmes concaténés par des tirets bas '_') ne se termine pas par un <é>, et se trouvant devant un mot commençant par une voyelle, à l'exception de <i> et <u>. La requête en (4.2) nous permet d'obtenir toutes les occurrences de noms ne se terminant pas par un <e>, optionnellement suivi de <s> ou <z>, et suivies par un mot commençant par une initiale vocalique hors <i> et <u>. L'ajout optionnel de <s> et <z> optionnels nous permet d'exclure les cas, nombreux, où c'est la flexion qui est détectée par Comparalem.¹ La requête (4.3) fait écho à la restriction sur les lemmes en (4.1) : elle nous permet d'obtenir toutes les occurrences de noms se terminant par un <e> mais dont le lemme se termine par un <é> — tous les cas où un <e> final en a.fr. représente /e/ ne sont pas couverts par des lemmes en -é, mais cette restriction permet déjà d'effectuer un tri, complété ensuite par le rejet manuel de certains lemmes. L'ajout de multiples signes dotés de diacritiques dans l'attribut lemma de cette requête répond au fait que les lemmes, parfois rédigés en français moderne, peuvent intégrer un ou plusieurs de ces signes composés, lesquelles n'appartiennent pas à la classe alphabétique [a-z] du langage CQP.

Les mêmes requêtes sont également effectuées dans le contexte d'une initiale consonantique (le second membre, `[word="(a|e|o|y).*"]`, est remplacé par 4.4) ainsi que sans approche contextuelle (le second membre est omis). Dans le cas de la requête comprenant l'initiale consonantique du mot suivant, le graphème <v> est omis puisqu'il provoque trop de faux-positifs (type *vne* pour *une*, cf. sec. 3.1.1), tandis que le graphème <h> est omis étant donné qu'il peut avoir un statut consonantique ou ne pas avoir de statut phonologique.² Les tables issues des requêtes (4.2-4.3) sont ensuite fusionnées en une seule table, représentant les formes qui ne se terminent pas par <e> pour /ə/, à l'aide de la commande cat en (4.5).

1: Sans cette précision, le programme détecterait par exemple *hunte* et *hunte*s comme des cas d'alternance (l'un se termine par <e> et l'autre pas); puisque l'argument est négatif (!=), nos données ne contiennent aucun mot qui se terminerait par un <s> ou un <z>, de manière à ne pas déséquilibrer les taux de défaut.

2: Sur quelques cas de 'h aspiré' dans nos données, cf. sec. 8.2.2.1 et 6.2.3.

(4.4) [...]_C :

```
[word="(b|c|d|f|g|j|k|l|m|n|p|q|r|s|t|w|x|z|ç).*"]
```

(4.5) cat NOM_nonE1_V.csv <(tail -n+2 NOM_nonE2_V.csv) > NOM_nonE_V.csv

Les six jeux de données ainsi obtenus sont représentés en (4.6). Les jeux NOM_nonE.csv, NOM_nonE_V.csv et NOM_nonE_C.csv résultent de la fusion (4.5) des résultats de requêtes du type de (4.2-4.3). Les jeux du type E comportent donc des formes qui se terminent par <e> pour /ə/, tandis que les jeux du type nonE comportent des formes ne se terminant pas par <e> pour /ə/, c'est-à-dire se terminant par <e> pour /e/ ou par un autre graphème vocalique.

(4.6) Jeux de données obtenus

NOM_E.csv	Occurrences se terminant par <e> pour /ə/;
NOM_nonE.csv	Occurrences ne se terminant pas par <e> pour /ə/;
NOM_E_V.csv	Occurrences se terminant par <e> pour /ə/ devant initiale vocalique;
NOM_E_C.csv	Occurrences se terminant par <e> pour /ə/ devant initiale consonantique;
NOM_nonE_V.csv	Occurrences ne se terminant pas par <e> pour /ə/ devant initiale vocalique;
NOM_nonE_C.csv	Occurrences ne se terminant pas par <e> pour /ə/ devant initiale consonantique.

4.1.1.2 Traitement des jeux de données dans Comparalem

Comparalem va donc, dans un premier temps, comparer les jeux de données NOM_E.csv et NOM_nonE.csv pour identifier les lemmes qui apparaissent dans les deux jeux. Dans un second temps, le programme va établir Tx_{\emptyset} , le taux de formes ne se terminant pas par <e> pour l'ensemble des ces lemmes pour chaque texte, c'est-à-dire le taux de défaut de <e> là où un <e> pour /ə/ serait attendu. L'établissement de ce taux est formalisé en (4.7). Le programme fait de même pour chaque contexte (Tx_{\emptyset_V} et Tx_{\emptyset_C} , devant initiale vocalique et consonantique). Ainsi, il permet d'étudier non pas le nombre d'occurrences sans <e> vis-à-vis du nombre absolu d'occurrences, mais uniquement vis-à-vis des occurrences des lemmes qui sont sujets à cette variation, ce qui permet d'observer la tendance de cette catégorie, sans qu'elle soit écrasée par les milliers ou dizaines de milliers d'occurrences de lemmes qui ne connaissent aucune variation. Lorsque l'ensemble d'une catégorie est interrogée, le taux de défaut absolu, portant sur tous les lemmes de la catégorie, est indiqué dans la colonne $Taux_{\emptyset/Nb}$. dans la table de présentation générale des résultats. La comparaison entre le Tx_{\emptyset} relatif et le $Taux_{\emptyset/Nb}$ absolu permet de mieux mesurer l'étendue de la variation.

- (4.7) Calcul du taux de formes à <e> pour /ə/
- Pour chaque lemme apparaissant à la fois dans NOM_E.csv et dans NOM_nonE.csv, et
 - Pour chaque texte,
 - Calculer Tx_{\emptyset} , le taux de défaut de <e> final :

$$Tx_{\emptyset} = \frac{nb.nonE}{nb.E + nb.nonE}$$

- Où $nb.E$ correspond au nombre d'occurrences de lemmes variables avec <e> pour /ə/, et
- Où $nb.nonE$ correspond au nombre d'occurrences de lemmes variables sans <e> pour /ə/.

Ce sont ces taux de défaut de <e> par texte qui forment le fondement principal de notre travail. Dans les applications cartographiques, ces taux par textes sont regroupés par région et les mesures (minimum, maximum, moyenne, écart-type) portent sur ces moyennes par région. La diffusion des défauts de <e> est représentée en fonction des lemmes dans les tables synthétiques et dans les graphiques qui représentent par un point non plus chaque texte mais chaque lemme de chaque texte; dans toutes les autres productions graphiques, les taux sont toujours ceux des textes.

4.1.2 Exclusions

En fonction des lemmes de Tree Tagger et des parties du discours de Dees remaniées par l'équipe du NCA, Comparalem a identifié 230 lemmes dotés à la fois de formes se terminant par <e> pour /ə/ et de formes ne se terminant pas par /ə/ pour schwa, soit un total de 54 590 occurrences. Une écrasante majorité de ces formes ne correspond pas à la cible de notre analyse et a donc été exclue. Cette exclusion est justifiée par le caractère morphologique de l'alternance <e>/∅ (4.8 a : c'est la déclinaison qui motive l'alternance), par l'existence d'un lemme unique pour une forme masculine et une forme féminine (4.8 b-c), par des erreurs d'annotation en parties du discours (4.8 d) ou en lemmes (4.8 e), par des ambiguïtés grapho-phonologiques (4.8 f), ou parce que l'irrégularité ne tient pas à l'absence d'un <e> mais à l'épenthèse d'un schwa (4.8 g) ou à une formation savante, éventuellement sous la forme d'un doublet (4.8 h).

(4.8) Causes d'exclusions et exemples

(a) Base variable

— « emperëor »³IMPERÁTOR → CSsg *emperere*IMPERATÓREM, -RES → CSpl/CRsg-pl *empeor(s)*

— « uem »

HÓMO → CSsg *uem*HÓMĪNEM, -NES → CSpl/CRsg-pl *home(s)*

(b) Nom masculin et nom féminin

— « amiet » : masc. *amiet*, fém. *amiète*

(c) Doublet nom masculin et féminin

— « plain » : masc. *plain*, fém. *plaine*— « uile » : masc. *oile*, fém. *oile*, *olie* ~ *olis*, *oli*

(d) Erreur d'annotation en POS

— « ceste » : ADJDÉM (et non NOM)

cist enfes qui tel mal endure

« Cette enfant qui souffre tant » (narcC, p. 22, trad.

pers.)

— « ceste » : PRODÉM (et non NOM)

Ipomedon ben s'aperceit ke ceste est en trop grant destreit

« Ipomedon s'aperçoit bien que c'est en trop grande

souffrance » (ipo, p. 184, *id.*)

(e) Erreur d'annotation en lemmes

— « braire » : confusion sous le même lemme de formes *brait*/*bret* issues de *braire* et de formes *braie* désignant le vêtement.*cri et brait* (athi, p. 52), *ne cri ne brait* (cambrai, p. 17)*chemise et braie tot ensamble* (chret1, p. 33; perr, p. 9)— « aigue » : toutes les formes sans <e> (ni <a> dans le Sud-Ouest) sont des formes du mot « œil » (p.ex. *iaulz*), pas des formes du mot « eau ».*ovre les iaulz et esgarde la grant lumiere* (barlaam, p. 103)

(f) <e> final pour /e/

— « gré2 » : *greit*, *gret*, *maugre* ~ *maugre*

(g) Forme épenthétique

— « or4_or » : *or*, *ore*Lorsque c'est bien le nom du métal, l'alternance *or* ~ *ore* tient à l'épenthèse d'un <e> pour /ə/; l'évolutionAURUM → *or* est attendue et majoritaire (788 formes sur 894).

(h) Formation savante

— « chalice » : *chalice*, *calice* ~ *chaliz*

Le premier type d'exclusions (4.8 a), dans lequel la variation est morphologique, tient principalement à l'alternance héritée du latin entre d'un côté CSsg paroxytons et de l'autre CSpl et CRsg-pl oxytons, du type *emperere(s)* ~ *empeor*, *larre(s)* ~ *larron*, *pute* ~ *putain(s)*, *none* ~ *nonain(s)*, etc. À ce type à déplacement d'accent s'ajoute le mot « uem » (fr.mod. « homme ») dont l'accent est fixe et dont c'est le CSsg qui est oxyton

3: Peu importe, pour l'utilisation de Comparalem et pour l'analyse synchronique de l'a.fr., que le mot soit de formation savante, ayant été ré-introduit dans le lexique après le couronnement de Charlemagne (É. BOURCIEZ et J. BOURCIEZ 1974 (1967), p. 40-41, Rq II); le mot est exclu puisque la déclinaison génère une alternance qui n'est pas un défaut de <e>.

narcC : ANON. (1165*). *Narcisus*. Ms. C; Dees nil; compo. norm., ca 1165; ms. pic., ca 1290; CR1 (PELAN et SPENCE 1964); DEAF : *NarcisusP*.

ipo : HUE DE ROTELANDE (ca 1185). *Ipomedon*. Compo. a.n., ca 1185; ms. a.n., ca 1190; CR (transcription inédite).

athi : ANON. (1200*). *Li romanz d'Athis et Prophilius*. Dees : Aisne; compo. pic., ca 1200; ms. frc., ca 1350; CR2 (HILKA 1912-1916); DEAF : *AthisH*.

cambrai : ANON. (1190*). *Raoul de Cambrai*. Dees : Ardennes; compo. Nord-Est, ca 1190; ms. pic., ca 1225; CR1 (MEYER et LONGNON 1882); DEAF : *RCambrM*.

chret1 : CHRÉTIEN DE TROYES (1180*). *Le Chevalier au lion & Perceval*. Dees : Hainaut; compo. champ. mérid., ca 1180; ms. champ., ca 1238; MS (transcr. inédite R. L. H. Lops); contenu présent ailleurs dans le corpus : incipit du *Chevalier au lion* (v. 1-998) dans yvf, et *Perceval* (v. 69-2000) dans perr.

perr : CHRÉTIEN DE TROYES (1180*). *Perceval*. Dees : [Hainaut]; compo. champ. mérid., ca 1180; ms. champ., ca 1237; MS1 (transcr. inédite R. L. H. Lops); inclus partiellement dans NCA. chret1.

barlaam : ANON. (1217*). *L'histoire de Barlaam et Josaphat*. Dees : Haute-Marne; compo. champ., ca 1217; ms. Est, ca 1290; CR3 (MILLS 1973); DEAF : *BaLJosPr¹M*.

CARLIER et al. (2020). « Catégories variables : noms, adjectifs, pronoms et déterminants ».

tandis que les autres cas sont paroxytons (CSsg *uem* ~ CSpl-sg-CRsg *home*) (CARLIER, GUILLOT-BARBANCE et al. 2020, GGHF, p. 637-639), à cause de processus de descendance paroxytonique de certains proparoxytons latins. L'autre nom de ce type, *cuens* (fr.mod. « conte »), n'a pas été détecté par Comparalem puisque Tree Tagger a visiblement échoué à placer le CSsg oxyton et les autres cas paroxytons sous le même lemme.⁴ Dans les données issues de Comparalem, ces deux types d'alternances morphologiques de patrons accentuels concernent 17 lemmes,⁵ soit 10 251 formes.

Par ailleurs, puisque nous n'avons pas filtré nos données en fonction du genre, Comparalem ne fait pas la différence entre les formes masculines et féminines, ce qui entraîne la détection d'une autre alternance <e>/Ø qui n'est pas phonologique mais morphologique.⁶ Cela concerne les noms dont le référent est animé et qui reproduisent le genre de celui-ci, comme le diminutif *amiet* et *amiète* (4.8 b), mais également des doublets au référent inanimé, comme *plain* et *plaine* (fr.mod. « plaine ») ou *oile*, *olie* (4.8 c). Dans l'exemple de *plain*, *plaine* c'est la forme masculine oxytone, issue du latin *PLĀNUS*, qui domine en a.fr., tandis que la forme féminine, issue d'une forme populaire †*plānia*, ne devient majoritaire que plus tard (TLFi; fréquence ici : 75 formes oxytones pour 7 formes paroxytones). Dans le cas de « uile », le mot peut être masculin ou féminin, mais les deux formes sont censées être paroxytones (FEW); nous n'avons que deux occurrences apparemment oxytones dans le NCA, *oli* (darm, p. 38) et *olis* (cambrai, p. 14). L'exclusion de ce mot est justifiée du fait la métathèse tardive de /j/ devant /ə/ posttonique (*uelie* → *oile*; SÉGÉRAL et SCHEER 2020, GGHF, p. 376-377); cette métathèse rend le type *oile* difficile à intégrer dans la même série que le type *olie* et ses versions défectueuses *oli*.

En (4.8 f), Comparalem détecte une alternance, mais celle-ci ne tient pas à l'alternance entre un patron oxyton et un patron paroxyton (donc à une alternance <e>/Ø) : le <e> final des 86 formes *malgre*, *maugre* correspond à la même voyelle phonologique /e/ que celui des 89 formes *gret*, *maugrez*. Le lemme a échappé à la restriction sur les lemmes se terminant par <é> (4.2), dans la mesure où l'étiquette de Tree Tagger se termine par un chiffre (gré2). On trouve également un certain nombre d'alternances <e>/Ø qui ne tiennent pas à un défaut de <e> mais à une épenthèse, comme *or* ~ *ore* (← *ÁURUM*) ou *doi* ~ *doie* (← *DÍGĪTRUM*). Un texte est particulièrement productif pour les épenthèses, et c'est sans surprise un texte anglo-normand : *edmund* est le seul à contenir les formes épenthétiques *curte* (p. 23, au lieu de *curt*), *morte* (7 occurrences,⁷ au lieu de *mort*) et *plure* (p. 17, au lieu de *plur*). Enfin, des formations savantes peuvent générer une forme canonique paroxytone, vis-à-vis de laquelle un défaut de <e> peut apparaître. C'est le cas de « chalice » (4.8 h) : la descendance de formation populaire du latin *CĀLIX*, *CĀLĪCEM* est *chaux* (« espèce de vase », FEW), et c'est donc par formation savante, dans le vocabulaire clérical, qu'on obtient les formes paroxytones *chalice* et *calice*. La forme en défaut de <e> qu'a détectée Comparalem, *chaliz*,⁸ représente bien le lemme clérical, et s'apparente donc à un défaut de <e>, mais n'étant pas prédictible à partir de l'étymon latin, elle est exclue de l'analyse. Sur les 230 lemmes détectés par Comparalem, ce sont donc 225 lemmes que nous excluons parce qu'ils correspondent à l'un des types donnés en (4.8).

4: Les formes du type *cuens* sont lemmatisées *conte1*. *conte* tandis que les formes du type *conte(s)* sont lemmatisées *conte*.

5: « emperëor, faitor, felon, gloton, jugëor, larron, lechëor, major, ome, paile, pastor, pechëor-peschëor, pejor, porteor, prodhomme (NCA pro),* traïtor, trichëor. »

* Les formes du mot *prodhomme* sont lemmatisées sous *pro* par Tree Tagger, et sont donc mélangées avec l'adj. *pro* (fr.mod. « preux »). Ce n'est pas le seul cas où le lemmatiseur se fonde uniquement sur une partie d'un nom composé (*damedies* lemmatisé *dan2*, *maugre* lemmatisé *gré2*, etc.).

6: Étant donné le faible nombre de résultats obtenus pour cette catégorie, nous n'avons pas restreint l'analyse par genre.

darm : ANON. (1275*). *Médecinaire Liègeois du xiii^e siècle*. Dees : Wallonie ; compo. liég., ca 1275 ; ms. wall, ca 1275 ; CR1 (HAUST 1941) ; DEAF : MédLiégH.

cambrai : ANON. (1190*). *Raoul de Cambrai*. Dees : Ardennes ; compo. Nord-Est, ca 1190 ; ms. pic., ca 1225 ; CR1 (MEYER et LONGNON 1882) ; DEAF : RCambRM.

7: P. 31, 43, 44, 45, 66 (x2), 74

8: *il eient beau chivals a mounter pore chaliz en le eglise e riche cope en servoise a manger, « ils avaient de beaux chevaux pour l'équitation, un pauvre calice dans l'église et une riche coupe pour le service du repas » (plainte, p. 6, trad. pers.)*

plainte : NICHOLAS BOZON (1312). *La plainte d'amour*. Dees : nil ; compo. a.n., 1312 ; ms. a.n., ca 1325 ; CR (BOZON 1905) ; DEAF : PlainteAmV.

4.2 Présentation des résultats

L'analyse de la variation <e>/Ø des noms ne porte que sur 5 lemmes, après exclusion des 225 lemmes ne correspondant pas, ou pas avec certitude, à la variation étudiée. Ces 5 lemmes sont « chacier » (subst. dérivé *chace*, fr.mod. *chasse*), « chataigne », « guerre », « novel » et « piece ». Ils représentent 884 occurrences, dont 6 seulement sont en défaut de <e> pour /ə/ final. Il faut donc noter, d'emblée, que très peu de lemmes à variation <e>/Ø (pour lesquels <e> pour /ə/ final est de droit) peuvent être détectés avec Comparalem.

Comme indiqué en table 4.1, ces 6 formes en défaut de <e> ne représentent qu'une infime partie de l'ensemble des lemmes de noms du corpus (0,07%) et des occurrences totales de tous ces lemmes ($1,4 \times 10^{-5}$, abrégé <0,01% sur la table). La façon dont cette table doit être lue est détaillée dans l'encadré qui la suit. La première colonne présente le nombre de lemmes annotés comme des noms (7467) et le nombre d'occurrences de ces lemmes (411 589). La seconde colonne présente le nombre de lemmes ayant une variation <e>/Ø finale que nous analysons comme un défaut de <e> (5 lemmes), et le nombre d'occurrences de ces lemmes (884). La troisième colonne présente la proportion de ces 5 lemmes sur l'ensemble des lemmes (0,07%) et de leurs 884 occurrences sur l'ensemble des occurrences (0,23%); ce taux renvoie à toutes les occurrences de ces lemmes, avec et sans défaut de <e>. La quatrième colonne présente le nombre d'occurrences en défaut de <e> (6 occurrences), et la cinquième colonne présente la proportion de ces occurrences défectueuses sur le nombre total d'occurrences des noms (<0,01%). Le taux de formes défectueuses de ces 5 lemmes (Tx_{\emptyset}), ici 6 formes sur 884 soit 0,68%, est représenté sur la dernière ligne (0,68%).

TABLE 4.1 – Quantité de noms, de lemmes à variation, et d'occurrences en défaut de <e>.

	Nb.	Nb.var.	Taux	Nb.Var.Ø	Taux.Ø/Nb.
Lemmes	7467	5	0,07%		
Occ.	411 589	884	0,23%	6	<0,01%
Tx_{\emptyset} (Nb.Var.Ø/Nb.var) :					0,68%

Comment lire la table 4.1

Le corpus contient 7467 lemmes annotés comme étant des noms, qui correspondent à 411 589 occurrences. Parmi ces 7467 lemmes, 5 sont dotés d'une variation <e>/Ø finale que nous avons analysée comme un défaut de <e>; ces 5 lemmes représentent 884 occurrences. Le taux de lemmes concernés par le défaut de <e> est donc de 0,07%, et le taux d'occurrences de ces lemmes (avec et sans <e>) est de 0,23%. Parmi ces 884 occurrences de ces 5 lemmes à variation <e>/Ø, 6 sont en défaut de <e> (<e> est présent dans 878 occurrences). Ces 6 occurrences représentent moins de 0,01% des 411 589 occurrences de noms du corpus ($Taux_{\emptyset}/Nb.$), et représentent 0,68% des 884 occurrences des 5 lemmes concernés (Tx_{\emptyset}).

Il est à noter que, dans cette table ainsi que dans les tables analogues produites pour les autres parties du discours, le nombre total de lemmes et de formes peut être sur- ou sous-évalué, dans la mesure où nous n'avons pas vérifié ces 7467 lemmes et leurs 411 589 occurrences. Or

Tree Tagger tend à multiplier les étiquettes lorsqu'il ne sait pas à quel lemme rattacher une forme, ce qui crée des lemmes composites (p.ex., *ici_or4_or* compte pour un lemme parmi les 7467 lemmes de noms, ce qui n'exclut en rien le fait que d'autres noms soient annotés *or4* ou *or*, ce qui ferait un total de trois lemmes au lieu de deux). À l'inverse, Tree Tagger renonce parfois à associer un lemme à une occurrence ; ainsi, 14 927 occurrences disposent de l'étiquette UNKNOWN pour l'attribut lemma. Si les lemmes UNKNOWN et leurs occurrences apparaissent bien dans les résultats absolus des tables récapitulatives, ils sont évidemment refusés en tant que témoins d'un défaut de <e>. Ainsi, en table 4.1 *supra*, UNKNOWN représente 1 lemme sur 7467, et près de 15 000 occurrences sur les 411 589 occurrences nominales relevées.

TABLE 4.2 – NCA : Lemmes de noms à défaut de Ø.

Lemme	Nb.	Nb.Ø	Tx_Ø
Guerre	529	2	0,4%
Piece	331	1	0,3%
Novel	15	1	6%
Chacier ^a	6	1	17%
Chastaigne	2	1	50%

a. Correspond au subst. *chace*, fr.mod. *chasse*.

La fréquence d'emploi de ces lemmes, représentée en table 4.2, va de 529 occurrences pour « guerre » (dont 2 en défaut de <e>) à 2 occurrences pour « chastaigne » (dont 1 en défaut de <e>). Seul le lemme *guerre* dispose de deux occurrences en défaut de <e>; les quatre autres n'en ont qu'une (cf. fig. E.2 et E.2 *infra*). Aussi, les cas de défaut de <e> sont toujours marginaux ; lorsque leur taux relatif au lemme est élevé, c'est uniquement parce que le lemme est, en lui-même, très rare dans le corpus.

4.2.1 Typologie

Puisque notre travail ne porte que sur les lemmes où /ə/ est de droit, les noms féminins apparaissent logiquement plus fréquemment que les noms masculins dans les cibles du défaut de <e>, au point où les 5 lemmes que nous avons retenus sont tous des substantifs féminins. Pour ces 5 lemmes, le schwa final est justifié par la présence d'un -A dans l'étymon latin (4.9 a), d'un /-a/ non latin mais qui suit le même traitement (4.9 b) ou par une formation déverbale féminine (4.9 c).

- (4.9) Origine du <e> pris en défaut
- (a) -A latin
— CASTÁNĚAM → *chastaigne*
— Lat.vulg. †*novella*⁹ → *nouvele*
- (b) /a/ non latin traité comme -A latin
— Frq. †*werra* → *guerre*
— Gaul. †*pettia* → *piece*
- (c) Dérivation au féminin
— *chace* dérivé de *chacier*

9: TLFi : « plur. neutre de *novellus* [...] pris comme subst. fém. singulier. »

L'occurrence en défaut de <e> du lemme « *piece* » dispose d'un statut particulier : elle appartient à la locution adverbiale *piece a*, qui renvoie à un temps passé et éloigné vis-à-vis de l'énonciation,¹⁰ parfois présentée en un mot (rendu *peça* par certaines éditions et dans certains dictionnaires), mais fréquemment présentée en deux mots dans les textes (AND), comme dans notre occurrence, reproduite en (4.10).

10: DMF « *piéça* » : « Il y a un certain temps, il y a longtemps, cela fait un certain temps »; FEW (« **pëttia* » 2.e§3) : « il y a longtemps, depuis longtemps », « naguère », « autrefois », etc.; AND « *peça* » : « some time ago; for some time; before, earlier on (in the text) ».

- (4.10) *traïtor estes tuit prové quant vos piec a ne l avez quis*
« Vous êtes traître, c'est certain, quand, il y a longtemps, vous l'avez ne l'avez pas cherché. »
(*cont ro*, p. 8, *id.*)

contro : ANON. (1200*). *La seconde continuation de Perceval*. Dees : Aube; compo. nil, ca 1200; ms. champ. mérid., ca 1225; CR1 (ROACH 1971); DEAF : *ContPerc*²R.

Cette locution adverbiale résulte de la fusion de *piece* dans son sens temporel¹¹ avec l'INDPRÉ3sg du verbe *avoir* (*pece a* : litt. « il y a du temps, il y a longtemps »; AND). Ainsi que le soulignent les graphies fusionnées *pieça*, *peça* etc., l'occurrence *piec* de *cont ro* n'a probablement pas le statut d'un substantif autonome, quand bien même la suppression du <e> dans la locution adverbiale n'est pas systématique.¹² Les cinq autres occurrences, en revanche, présentent une autonomie nominale normale.

11: FEW (« **pëttia* » 2.e§1) : *une pece* « quelque temps »; *une piece del tans* [litt. « une division du temps »]; *grant piece est* « il y a longtemps » (insertion entre crochets de notre fait).

12: P.ex. *le vostre fil piece a que il est mors*, « Votre fils, qui est mort depuis longtemps » (*orso*, p. 55, trad. pers.).

orso : ANON. (1225*). *Orson de Beauvais*. Dees : Marne; compo. pic. mérid., ca 1225; ms. lorr., ca 1290; MS2 (ANON. 1899); DEAF : *OrsonP*.

4.2.2 Distributions

Comme indiqué *supra*, non seulement les lemmes en défaut de <e> que Comparalem a permis d'identifier sont rares, mais leurs défauts de <e> sont également marginaux. Les données sont ici trop faibles pour qu'un approche statistique soit valable, aussi l'étude des distributions diachroniques et diatopiques de ces formes en défaut de <e> se résume à la description des occurrences en question. Nous ne suivons donc pas la structure qui sera déployée pour les autres parties du discours, et certaines figures sont rejetées en annexe (p. 611 *et sq.*) plutôt qu'insérées dans le corps du texte.

Les deux occurrences *guer* se trouvent dans *darm*, au sein d'un extrait qui accumule, en parataxe et de manière formulaire, des explications de rêves (4.11, 4.12). Le texte, wallon avec un fort degré de précision (coef. dial. : 94), est daté de 1275. L'occurrence *piec*, citée *supra* en (4.10), apparaît dans *cont ro*, un texte daté du début du XIII^e s. et que Dees localise dans l'Aube, encore une fois avec un fort degré de précision (coef. dial. : 91). L'occurrence *novel*, dont le genre féminin est confirmé par le déterminant *une*, apparaît dans *chauvency* (4.13), un texte de Haute-Marne composé à la fin du XIII^e s. et copié au début du siècle suivant. L'occurrence *chaz*

pour *chace* est peu claire en contexte, mais son placement au côté de *chiens* et *lous* dans la coordination en *ne* semble bien défendre son statut de substantif (4.14); elle apparaît dans *cont re*, la sixième branche du *Roman de Renart*, texte daté du début du xiv^e s. pour la composition comme pour la copie, et localisé dans l'Aube par Dees. Enfin, l'occurrence *chastain* apparaît dans *edmond* (4.15), hagiographie anglo-normande composée au milieu du xiii^e s. et copiée à la fin de ce siècle; le mot y prend son sens figuré de « objet de peu de valeur » (AND, FEW « castanea » 1.a§2), dans un passage de critique de la vanité de la vie de la noblesse.

- (4.11) *Lion veir corochier guer de son anemis senefie.*
« Voir un lion courroucé signifie une guerre de la part de son ennemi. »
(*darm*, p. 12-13, trad. pers.)
- (4.12) *Cui hurs assailhent, guer de ses anemis senefie.*
« Qui se fait attaquer par des ours, [cela] signifie une guerre de la part de ses ennemis. »
(*darm*, p. 14, *id.*)
- (4.13) *je conte tout la novel au roi et a son chevalier*
« Je raconte toute l'histoire au roi et à son chevalier »
(*chauency*, p. 3, *id.*)
- (4.14) *onques ne chiens ne chaz ne lous plus mauves morcel ne manja*
« Jamais ni chien, ni [animal de] chasse, ni loup ne mangea quelque chose de plus mauvais »
(*cont re*, p. 35, *id.*)
- (4.15) *ne les prise a une chastain teus est li mund*
« ne leur donne pas plus de prix qu'à une châtaigne. Ainsi est le monde... »
(*edmond*, p. 24, *id.*)

Ainsi, l'occurrence la plus précoce est *piec*, au début du xiii^e s., tandis que l'occurrence suivante, *chastain*, n'apparaît qu'un demi siècle plus tard. Comme illustré en fig. 4.1, page suivante, les quelques occurrences à défaut de <e> se concentrent donc dans la seconde moitié du corpus, et, qui plus est, la première occurrence est exceptionnelle dans la mesure où *piec* dans *piec a* peut ne représenter que la lexicalisation de la locution *piece a*. Ces quelques occurrences, si elles doivent être considérées comme représentatives d'une tendance phonologique, sont donc tardives.

darm : ANON. (1275*). *Médecinaire Liégeois du xiii^e siècle*. Dees : Wallonie; compo. liég., ca 1275; ms. wall, ca 1275; CR1 (HAUST 1941); DEAF : MédLiégH.

chauency : JACQUES BRETTEL (1285*). *Le tournoi de Chauency*. Ms. M; Dees : Haute-Marne; compo. Meuse, ca 1285; ms. lorr., ca 1310; CR1 (JACQUES BRETTEL 1932); DEAF : BretTournD.

cont re : ANON. (1322*). *Le Roman de Renart le contrefait, fin de la 6^e et dernière branche*. Dees : Aube; compo. champ. mérid., ca 1322; ms. nil, ca 1337; CR2 (LEMAÎTRE et RAYNAUD 1914); DEAF : RenContr¹R

edmond : MATTHIEU PARIS (1255*). *La vie de saint Edmond*. Dees : nil; compo. a.n., ca 1255; ms. a.n., ca 1280; MS2 (J.-A. BAKER 1929); DEAF : SEdmCantB

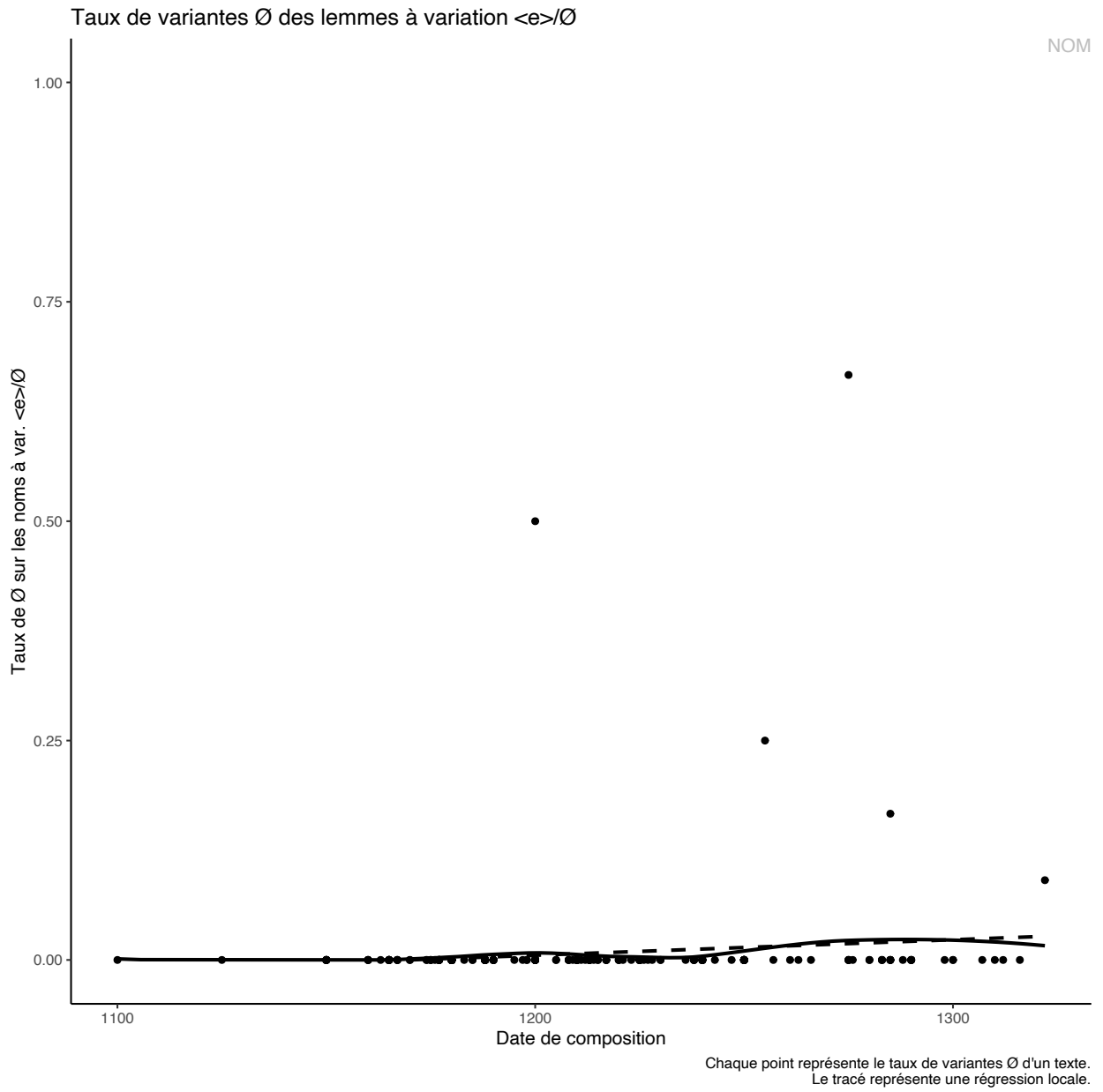
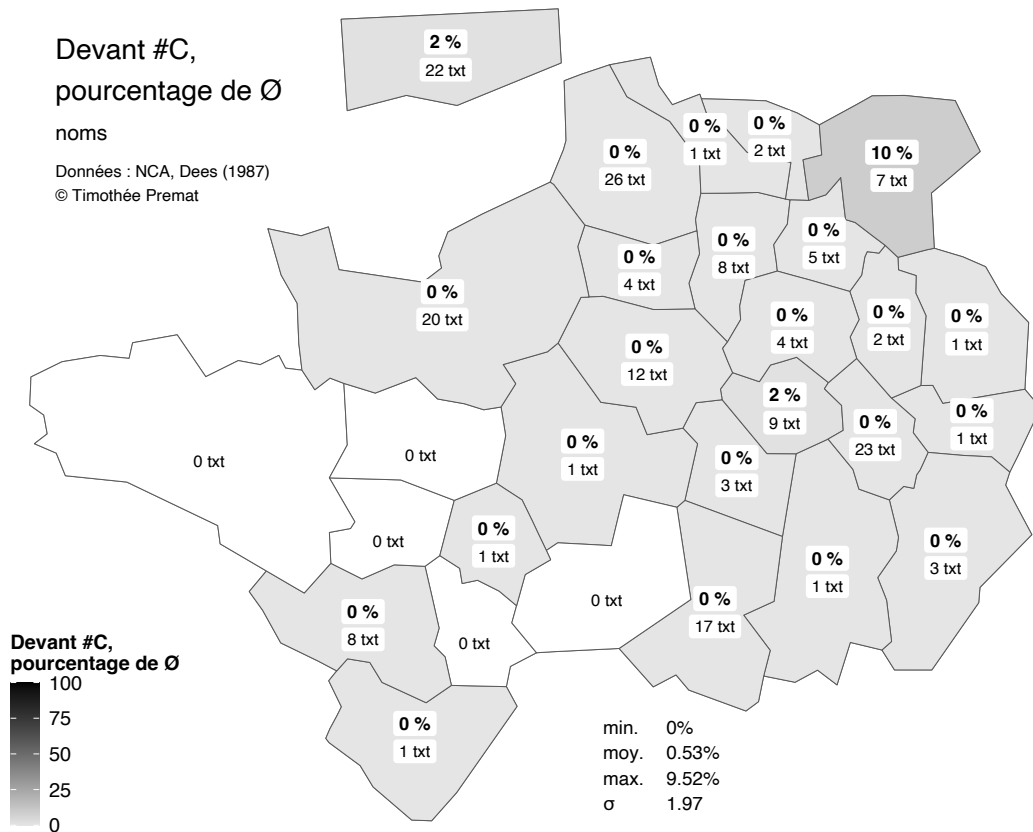


FIGURE 4.1 – Distribution diachronique des variantes de noms en défaut de <e>.

Du point de vue diatopique, 5 occurrences en défaut de <e> sur 6 se concentrent dans le quart Nord-Est du domaine : les deux occurrences *guer* apparaissent dans un texte wallon, *piec* et *chaz* apparaissent dans l'Aube, *novel* apparaît dans la Haute-Marne ; seule l'occurrence *chastain* ne provient pas du Nord-Est, apparaissant en anglo-normand. Parallèlement, l'absence de défaut de <e> est attesté dans les autres régions, principalement grâce aux lemmes « guerre » et « piece », comme le représentent les cartes en fig. 4.2, page suivante. Les taux positifs, sur ces cartes, sont à manier avec précaution étant donné le très faible nombre de lemmes interrogés. Pour la même raison, nous avons exceptionnellement désactivé le filtre qui écarte les régions contenant moins de 5 textes. On constate néanmoins qu'approximativement les deux tiers Sud-Ouest du domaine, c'est-à-dire tout le territoire sauf le Nord-Est et l'Angleterre, ne témoignent d'aucune occurrence de défaut de <e> pour les noms, alors que ces régions disposent souvent d'occurrences d'au moins un des cinq lemmes en question, et même régulièrement de plusieurs occurrences d'un ou plusieurs de ces lemmes (principalement « guerre » et « piece »).

Enfin, le contexte subséquent ne semble pas jouer un rôle déterminant pour ces cinq occurrences : quatre sont devant initiale consonantique (4.11, 4.12, 4.14, 4.15), et deux devant initiale vocalique (4.10, 4.13). Si l'on ajoute à cela que, parmi les occurrences placées devant une initiale vocalique, se trouve l'occurrence *piec a*, il faut admettre que, dans ces données minimales, une initiale vocalique ne peut être considérée comme favorisant un défaut de <e> pour les noms. Même sans écarter *piec a*, le test de Student en fig. 4.3 est non significatif ($p = 0,6$ ou $0,4$) et l'effet mesuré est faible ($t = 0,49$). Ici, expliquer les occurrences *piec a* et *novel au* par leur contexte subséquent serait donc un non-sens au vu du reste des données.

a



b

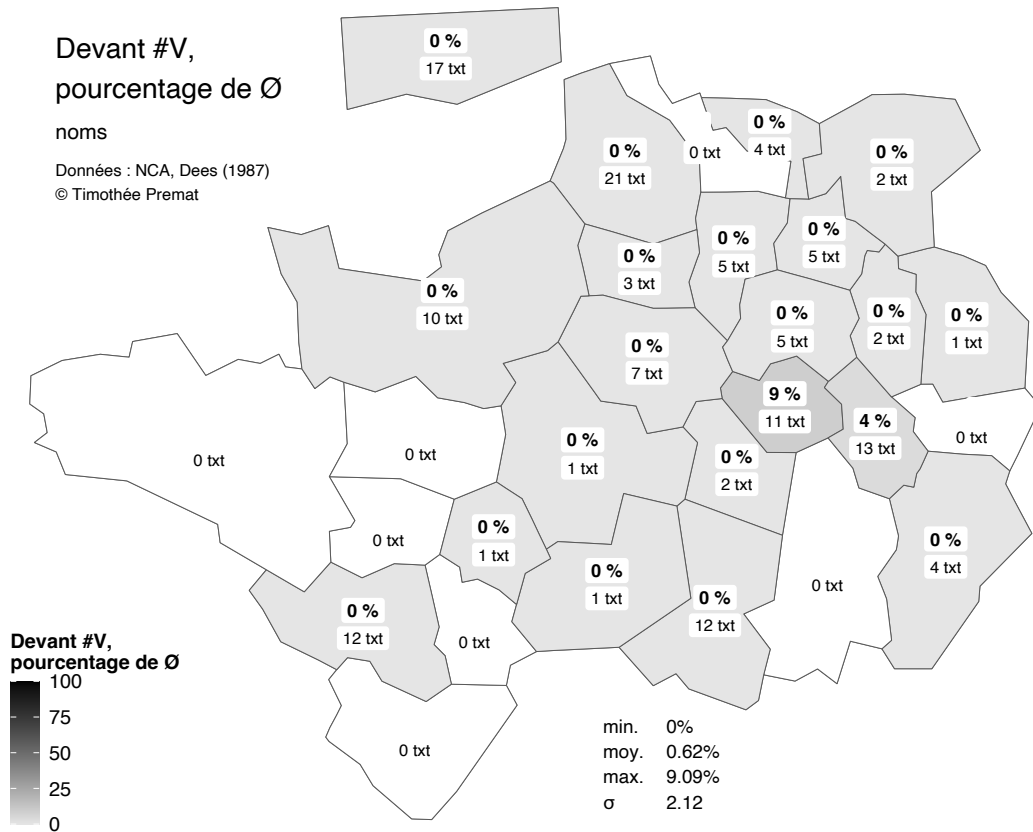


FIGURE 4.2 – Distribution diatopique des variantes de noms en défaut de <e>.

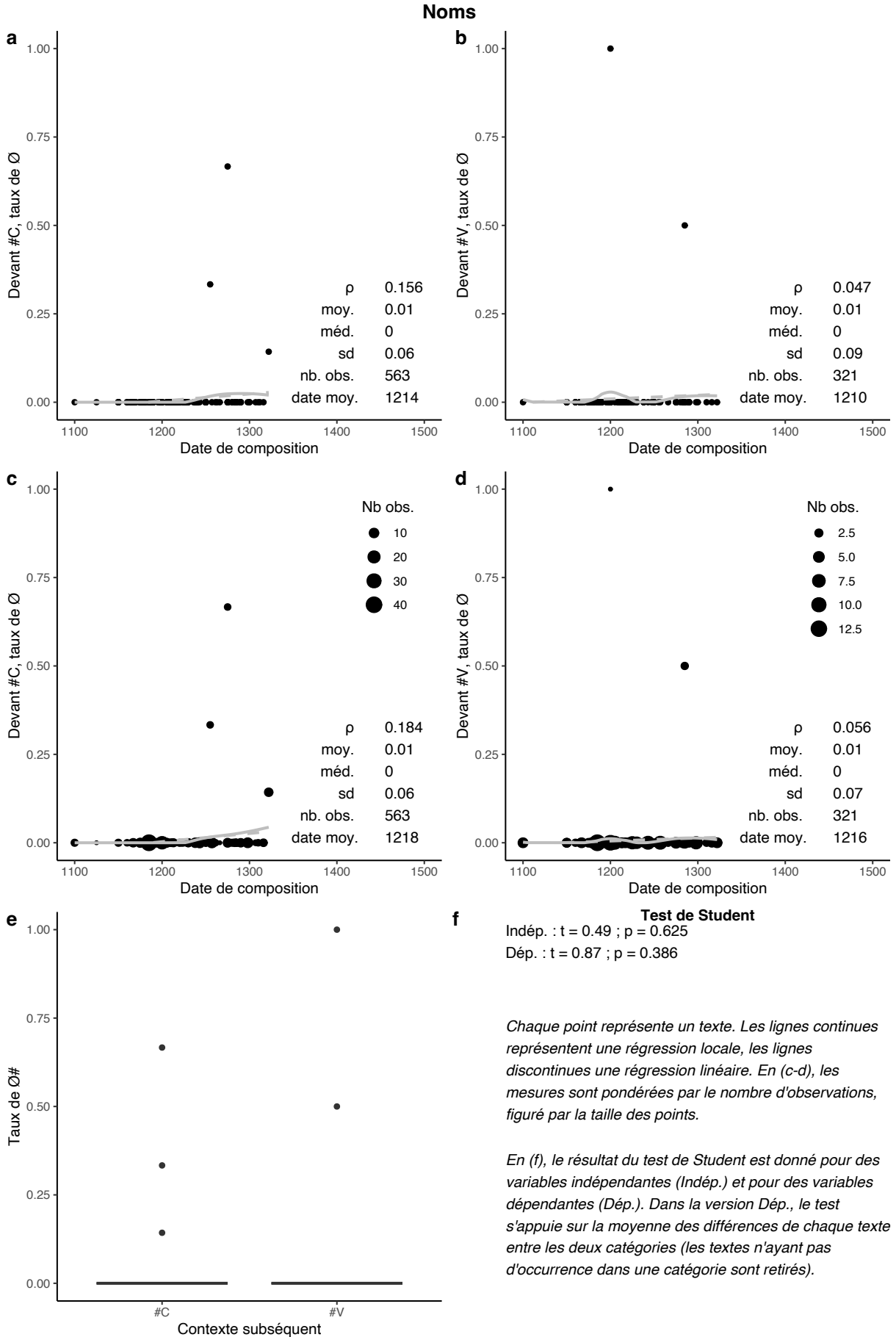


FIGURE 4.3 – Distribution contextuelle des variantes de noms en défaut de <e>.

4.3 Généralisations

Comparalem n'a pas permis la mise en valeur d'une variation d'ampleur dans la présence d'un <e> pour /ə/ final dans la classe des noms. Cela peut avoir deux origines : soit le programme et/ou le corpus et ses annotations n'ont pas permis la détection d'une variation qui, pourtant, existe bien dans les textes, soit cette variation n'existe presque pas dans les textes. Les deux facteurs peuvent être simultanément en jeu, une variation rare touchant des lemmes rares mettant en difficulté le lemmatiseur. Un travail complémentaire, présenté en fin de partie, et effectué sur le corpus à lemmatisation vérifiée de la BFM (version BFM3) confirme l'extrême marginalité des cas de défaut de <e>, mais témoigne d'occurrences légèrement plus nombreuses.

Nous ne pouvons évidemment pas proposer de généralisations positives sur le fondement de six occurrences de cinq lemmes. Néanmoins, nous pouvons proposer des généralisations négatives, quoi que limitées par la méthode : sur les noms dont le schwa final est de droit, Comparalem n'a permis la détection que d'un nombre très restreint de lemmes sujets à défaut de <e> (0,07% des lemmes), et dont les formes en défaut de <e> sont marginales, cette marginalité étant proportionnelle au nombre brut d'attestations du lemme (de 0,4 et 0,3% des occurrences pour « guerre » et « piece » à 6, 17 et 50% des occurrences pour « novel », « chacier » (« chace ») et « chastaigne » ; <0,01% des occurrences totales).

Si ces occurrences n'apparaissent pas dans la première partie du corpus (*piec a* : ca 1200 ; *chastain* : ca 1255 ; les autres ensuite), le nombre d'occurrences détectées est trop faible pour soutenir une déduction diachronique un tant soit peu sérieuse. Par ailleurs, s'il faut remarquer que 5 occurrences apparaissent dans le quart Nord-Est et une en anglo-normand, et que le reste du domaine ne témoigne d'aucun défaut de <e> pour les noms, cette conclusion doit également être lourdement nuancée par le nombre trop restreint d'occurrences, et par le fait que, même au sein du Nord-Est et de l'aire anglo-normande, les taux de défaut de <e> demeurent marginaux. De plus, il faut remarquer que deux des six occurrences proviennent de *dar*m, dont le style formulaire et parataxique peut indiquer un statut (socio-)linguistique spécifique ; ce texte se départ des autres textes, narratifs, qui attestent des quatre autres occurrences. C'est aussi dans *dar*m que se trouve l'une des deux occurrences en défaut de <e> du lemme « uile » (*oli*, p. 38). Il se pourrait donc que ce texte, pour des facteurs qui tiennent à sa nature textuelle, ne soit pas un témoin comme les autres et ne puisse guère être mélangé avec l'ensemble des textes.

Enfin, du point de vue du contexte subséquent, un aussi faible nombre d'occurrences ne saurait soutenir une analyse sérieuse. Néanmoins, même au sein de cet ensemble restreint de lemmes et d'occurrences, il faut remarquer que le contexte antévocalique ne semble pas favoriser le défaut de <e>.¹³

13: Cf. fig. E.4, en annexes (p. 613), pour une vue complémentaire du croisement entre contexte subséquent et diatopie.

4.4 Conclusion

Si des observations ponctuelles peuvent être menées sur ces occurrences (poids et statut de *darm*, chronologie et diatopie), celles-ci demeurent bien en deçà d'un quelconque seuil numérique qui assurerait leur intérêt et leur fiabilité dans une analyse diachronique et diatopique. Aussi, la conclusion de ce chapitre ne peut tenir qu'à remarquer que *Comparalem* n'a pas permis, sur le corpus du NCA, de mettre en lumière une réelle variation <e>/Ø en position posttonique finale pour les noms.

Adjectifs au féminin singulier

5.1 Introduction

Après avoir interrogé les défauts de <e> dans les noms, nous nous intéressons à une autre catégorie lexicale, celle des adjectifs. Nous avons restreint notre analyse aux adjectifs féminins singuliers. Dans ce chapitre, après avoir justifié ce choix de la restriction aux adjectifs féminins singuliers (sec. 5.1.1), avoir couvert les exclusions de lemmes et de formes (sec. 5.1.2) et les quelques fusions de lemmes que nous avons opérées (sec. 5.1.3), nous présentons les résultats concernant les défauts de <e> (sec. 5.2), avant de discuter ces résultats pour les doter d'une portée analytique et monter en généralité (sec. 5.3).

La description des données (sec. 5.2) présente une brève typologie des lemmes à défaut de <e> (sec. 5.2.1), avant d'aborder différentes variables et leur rôle dans la distribution des défauts de <e> : en fonction de la diachronie (sec. 5.2.2), en fonction des lemmes et de leur fréquence (sec. 5.2.3), en fonction du contexte subséquent (sec. 5.2.4), en fonction de la diatopie (sec. 5.2.5), et en fonction de l'interface entre diatopie, diachronie et contexte subséquent (sec. 5.2.6). La phase analytique et généralisante (sec. 5.3) extrait les éléments pertinents de l'analyse des variables que représentent les lemmes et leur fréquence (sec. 5.3.1), le contexte (sec. 5.3.2), la diachronie (sec. 5.3.3) et la diatopie (sec. 5.3.4), avant de développer une analyse plus poussée de la place et du poids de l'anglo-normand dans la diachronie et dans la diatopie du corpus (sec. 5.3.5), pour conclure en prenant du recul sur la place des adjectifs féminins singuliers à défaut de <e> au sein de l'ensemble plus large des adjectifs féminins (sec. 5.3.6). Cette section propose des généralisations formelles qui, tout en grossissant le grain de l'analyse, permettront la comparaison avec les autres catégories grammaticales analysées.

5.1.1 Choix des adjectifs au féminin singulier

Étant donné que la présence d'une consonne finale ne permet pas le déclenchement de l'élision, nous avons exclu les adjectifs masculins, dont la flexion à accentuation régulière impose un -s au CSsg et au CRpl (5.1), et les adjectifs féminins, dont la flexion à accentuation régulière impose un et -s au pluriel (5.2).

- | | | | | | |
|-------|-------|------|------------------|------|------------------|
| (5.1) | Masc. | CSsg | <i>estranges</i> | CSpl | <i>estrange</i> |
| | | CRsg | <i>estrange</i> | CRpl | <i>estranges</i> |
| (5.2) | Fém. | CSsg | <i>estrange</i> | CSpl | <i>estranges</i> |
| | | CRsg | <i>estrange</i> | CRpl | <i>estranges</i> |
- (MOIGNET 2002 [1976], p. 27)

MOIGNET (2002 [1976]). *Grammaire de l'ancien français*.

Cette exclusion des adjectifs masculins et des adjectifs féminins pluriels n'est pas problématique du point de vue quantitatif, dans la mesure où,

une fois les exclusions ponctuelles appliquées également au niveau des lemmes et des occurrences, notre corpus correspond à 40 228 occurrences d'adjectifs féminins singuliers. Ainsi, nous considérons que le corpus sur lequel nous travaillons ici est suffisamment grand pour être considéré représentatif, malgré l'exclusion des adjectifs masculins et des adjectifs féminins pluriels. Par ailleurs, ce corpus est aussi plus simple, grammaticalement parlant : inclure les adjectifs masculins signifierait inclure également des formes fautives sur le plan de l'accord, ou des formes dont la base se termine en *-s*, rendant invisible l'accord et compliquant encore la situation. On échappe ainsi aux formes de masc. CSsg en défaut de *-s* (5.3 - 5.5) et aux formes de masc. CSpl où *-s* s'est étendu ou fait partie de la base (5.6 - 5.7); *-s* commence à fonctionner comme un marqueur de pluriel et non plus de cas, à mesure que la déclinaison s'affaiblit, voire disparaît dans l'Ouest (CARLIER, GUILLOT-BARBANCE et al. 2020, GGHF, 645-648); sa régularité fait défaut en anglo-normand dès le XII^e s. (SHORT 2007, p. 122-123).

- (5.3) *Aucassins fu lie.*
« Aucassin fu réjoui. »
(auc, p. 7, trad. pers.)
- (5.4) *molt estoit li nains debonaire*
« Le nain était très généreux »
(bel, p. 11, id.)
- (5.5) *biau sire*
« cher seigneur »
(aye, p. 3, id.)
- (5.6) *tant chivoaigerent li corteis mesager que...*
« Les courtois messagers chevauchèrent tant que... »
(alia, p. 16, id.)
- (5.7) *li religious moyne qui hardemment amoient nostre segnor*
« Les dévots moines qui aimaient hardiment notre seigneur »
(barlaam, p. 3, id.)

Ainsi, en (5.3 - 5.5), le cas sujet est non seulement requis par la syntaxe, mais il est également attesté sur le nom auquel se rapporte l'adjectif : *-s* sur *Aucassins* (5.3) et sur *nains* (5.4); forme paroxytone pour *sire* en (5.5) (*seignor(s)* aux autres cas) et, en (5.4), par le déterminant *li*. De même au CSpl masculin : en (5.6 - 5.7), le substantif est correctement accordé au CSpl (*li* [...] *mesager*, *li* [...] *moyne*), alors que l'adjectif porte un *-s* qui n'est pas dû à sa flexion mais appartient à sa base (*corteis*, *religious*). De fait, ce type d'occurrences provoque un bruit dans les données. Se concentrer sur les formes de féminin singulier permet donc également d'exclure le bruit que générerait cette variation. Au demeurant, savoir ce qui tient du phénomène phonologique ou morphologique dans ces occurrences est une question difficile, alors que la présence ou l'absence d'un *-s* final a une incidence phonologique claire : le schwa qui précéderait un tel *-s* ou défaut de *-s* devient ou cesse d'être final (p.ex. *debonaire* au lieu de *debonaires*).

CARLIER et al. (2020). « Catégories variables : noms, adjectifs, pronoms et déterminants ».

SHORT (2007). *Manual of Anglo-Norman*.

auc : ANON. (1225*). *Aucassin et Nicolette*. Dees : Pas-de-Calais; compo. pic., ca 1225; ms. pic., ca 1290; MS1 (ROQUES 1955); DEAF : AucR³; ne contient que les parties en prose.

bel : RENAUT DE BEAUJEU (1200*). *Le bel inconnu*. Dees : [Somme, Pas-de-Calais]; compo. Est, ca 1200; ms. hain., ca 1283; MS2 (RENAUT DE BEAUJEU 1929); DEAF : RenBeaujBeLW².

aye : ANON. (1200*). *Orson de Beauvais*. Dees : [Normandie]; compo. pic./norm., ca 1200; ms. nil, ca 1300; CR1 (ANON. 1967); DEAF : AyeB.

alia : ANON. (1185*). *Roman d'Alexandre*. Ms. A; Dees : Vendée, Deux-Sèvres; compo. pic. et occ. orient., ca 1185; ms. poit., ca 1225; MS (ANON. 1937); DEAF : AlexArsL.

barlaam : ANON. (1217*). *L'histoire de Barlaam et Josaphat*. Dees : Haute-Marne; compo. champ., ca 1217; ms. Est, ca 1290; CR3 (MILLS 1973); DEAF : BalJosPr^{1M}.

5.1.2 Exclusions

Comme pour toutes les parties du discours, nous avons utilisé les deux entrées d'exclusions de Comparalem, en sélectionnant uniquement les lemmes pour lesquels <e> est de droit et qui ne comportent pas majoritairement des erreurs d'annotation (ce qui provoque l'exclusion de tous les autres lemmes), ainsi qu'en dressant la liste des formes pour lesquelles le défaut de <e> n'est pas avéré.

Comparalem a détecté 108 lemmes sujets à défaut de <e>; nous en avons conservé 45. Le motif d'exclusion le plus courant est la présence d'adjectifs épiciques. Pour ces adjectifs, la forme attendue est oxytone, étant donnée l'absence de marque de genre en latin, qui se perpétue en ancien français (5.8).

(5.8)	Cas	Étymon		Masc.	Fém.
	CSsg	GRÁNDIS	→	granz	granz ou grant
	CRsg	GRÁNDEM	→	grant	grant
	CSpl	GRÁNDES	→	grant	granz
	CRpl	GRÁNDES	→	granz	granz

(Formes issues de MOIGNET (2002 [1976], p. 28))

Dans le cas de ces adjectifs, c'est la présence d'un <e> final au féminin qui est analogique et qui justifie l'exclusion : l'opposition *grande* ~ *grant* s'analyse en termes d'ajout *-e* analogie sur le modèle des adjectifs féminins réguliers, où *-e* marque le féminin, non en termes de défaut d'un <e> dans *grant*. Les lemmes concernés sont au nombre de 16; ils ont été exclus sur le seul critère de la déclinaison latine de leur étymon.¹ Le deuxième motif d'exclusion le plus répandu est la déclinaison à base variable, qui fonctionne ici comme pour les noms (cf. sec. 4.1.2). Dans notre corpus, ce type comprend des comparatifs (*graindre*, *graignor*; *maire*, *major*; *mieudre*, *meillor*; *mendre*, *menor*; *pire*, *pejor*) et le substantif adjectivé *fel*, *felon*. Dans le cas des comparatifs, le féminin ne se distingue en principe pas du masculin, et seul le CSsg dispose d'un <e> (5.9). Dans le cas de *fel*, *felon*, le type féminin *felonnesse* domine (28 occ. sur 30), que ce soit au CS ou au CR, mais deux occurrences *fel* issues du CS induisent la détection de la variation <e>/Ø par Comparalem.

(5.9)	Cas	Étymon		Masc.	Fém.
	CSsg	MÉLŶOR	→	<i>mieudre(s)</i>	<i>mieudre</i>
	CRsg	MELIÓREM	→	<i>meillor</i>	<i>meillor</i>
	CSpl	MELIÓRES	→	<i>meillor</i>	<i>meillors</i>
	CRpl	MELIÓRES	→	<i>meillors</i>	<i>meillors</i>

(Formes issues de MOIGNET (2002 [1976], p. 30))

Enfin, nous avons exclu les lemmes dont toutes les occurrences en défaut de <e> n'étaient pas fiables ou étaient fausses; c'est par exemple le cas chaque fois qu'un adjectif se rapporte à *amour* ou à *gent* (qui peuvent avoir les deux genres en a.fr.), ou chaque fois qu'un adjectif a été annoté comme féminin alors qu'il est masculin.² Lorsque l'exclusion ne concerne que certaines formes en défaut de <e> d'un lemme, ces formes ont été introduites dans le dictionnaire d'occurrences à exclure, principalement à cause de graphies <a> dans le Sud-Ouest,³ ou à cause d'erreurs ponctuelles d'annotation (en parties du discours ou en lemme),⁴ de

1: « autretel, brief, criuel, criuos, fort, grant, grief, itel, léal, pro*, quel, quell, quell_ quel, quell_ quel_ quelque, tel, vert. »
*Pour les adjectifs, le lemme « pro » correspond au fr.mod. *preux*.

2: P.ex. f1o, p. 10 : *et romain sont de guerre hardi et alosé*, « et les romains sont hardis à la guerre et renommés [pour cela] »; l'annotateur a probablement été trompé par la précession immédiate de *guerre*, interposé entre le verbe et l'attribut.

3: P.ex. 5 occ. *bona* dans *poit*.
Les <a> finaux sont détectés par Comparalem comme des défauts de <e> (la forme ne se termine pas par <e>; cf. sec. 3.2.1, p. 71).

poit : ANON. (1217*). *Chronique traduite de l'Historia Karoli Magni et Rotholandi*. Dees : Vendée, Deux-Sèvres; compo. saint., ca 1217; ms. saint., ca 1238; MS1 (AURACHER 1877); DEAF : Turpin¹A.

4: P.ex. edmund, p. 36 : *reisun dreit ne mesure n a*, « il n'a ni raison, ni droit, ni mesure » : *dreit* n'est probablement pas Adj de *reisun*, mais est annoté comme tel.

edmund : DENIS PIRAMUS (1195*). *La Vie Seint Edmund le Rei*. Dees : [Angleterre]; compo. a.n., ca 1195; ms. a.n., ca 1325; MS2 (DENIS PIRAMUS 1935); DEAF : EdmK.

rapports ponctuels avec un nom dont on ne peut établir le genre (p.ex. *amour, gent*), ou parce qu'elles apparaissent dans des passages obscurs qui ne nous permettaient pas de garantir à la fois l'annotation de l'adjectif et le genre du nom auquel l'adjectif se rapporte. Ce sont ainsi 114 occurrences qui ont été exclues. Une fois ces exclusions de lemmes et de formes appliquées, les 45 lemmes conservés correspondent à 9374 occurrences, dont 228 sont en défaut de <e> (2,43%).

5.1.3 Relemmatisation

Face aux résultats préliminaires, et en ayant vérifié toutes les occurrences en défaut de <e> concernées, nous avons procédé à une relemmatisation de certains lemmes, pour fusionner des lemmes que Tree Tagger avait séparés alors qu'il apparaît, à la lecture des occurrences, qu'il s'agit bien du même lemme. Les lemmes fusionnés sont indiqués en (5.10); pour chaque item de cette liste, toutes les possibilités de lemmes composés sont prises en compte (ainsi, *gent_relemma* représente la fusion de *gent*, *gent2*, *gent_gent2* et *gent2_gent*).

(5.10)	Tree Tagger	Nouveau lemme
	<i>coart, cöart, coärt</i>	→ <i>coart_relemma</i>
	<i>fol, fol1, fol2, fol3</i>	→ <i>fol_relemma</i>
	<i>gent, gent2</i>	→ <i>gent_relemma</i>
	<i>mal, mall</i>	→ <i>mal_relemma</i>
	<i>prest, prest1</i>	→ <i>prest_relemma</i>
	<i>sol, sol1</i>	→ <i>sol_relemma</i>
	<i>sor, sor1, sor2</i>	→ <i>sor_relemma</i>
	<i>vair1, voir, vrai</i>	→ <i>vair_relemma</i>

5.2 Présentation des résultats

Les 45 lemmes sélectionnés représentent un peu plus de 5% des lemmes dotés de formes annotées en tant qu'adjectifs féminins singuliers, comme indiqué en table 5.1. En termes d'occurrences, ces 45 lemmes représentent 9374 formes, dont 228 seulement sont en défaut de <e>. Certains de ces 45 lemmes sont parmi les plus fréquents, puisque un peu plus de 5% des lemmes représentent un peu plus de 23% des occurrences d'adjectifs féminins singuliers. Ces 228 occurrences représentent un peu plus de 2,4% des 9374 occurrences de ces lemmes, et un peu moins de 0,6% de toutes les occurrences d'adjectifs féminins singuliers. La variation <e>/Ø ici analysée ne concerne donc qu'une petite partie (2,4%) d'un sous-ensemble (23% des occurrences, 5% des lemmes) au sein des adjectifs féminins singuliers (soit 0,6% des occurrences).

TABLE 5.1 – Quantité d'adjectifs féminins singuliers, de lemmes à variation, et d'occurrences en défaut de <e>.

	Nb.	Nb.var.	Taux	Nb.Var.Ø	Taux.Ø/Nb.
Lemmes	837	45	5,38%		
Occ.	40 228	9374	23,3%	228	0,57%
		<i>Tx_Ø (Nb.Var.Ø/Nb.var) :</i>			2,43%

5.2.1 Typologie

Du fait du filtre ne retenant que les adjectifs féminins singuliers, la typologie grammaticale des formes à variation est assez simple : ce sont tous des adjectifs à accentuation régulière au féminin singulier, cas sujet ou cas régime. Du point de vue phonologique, seuls trois lemmes ont une base qui se termine par une voyelle : « corsu » (fr.mod. « corpulent » (5.11)), « piu » (fr.mod. « pieux » (5.11)) et « vrai » (fr.mod. « vrai » (5.13)).

- (5.11) *la tres pieu dame lour dist*
« la très pieuse dame leur dit »
(jaco, p. 20, trad. pers.)
- (5.12) *la reine fu corsu e grant*
« la reine était plantureuse et grande »
(fablesA, p. 4; fablesB, p. 3; fablesC, p. 4, id.)⁵
- (5.13) *verai savoir verai vertu verai avoir veraie vertu*
« Savoir véritable, vertu véritable, possession véritable, vertu véritable »
(calex, p. 10, trad. pers.)

La forme féminine attendue pour ces lemmes comprend un hiatus V_o final : *corsue*, *pieue* et *veraie*. Ces formes hiatiques sont d'ailleurs attestées dans d'autres textes dans le même contexte; on trouve ainsi le syntagme *pieue dame* dans *mir* (5.14), tandis que d'autres versions des *Fables* de Marie de France ont bien *corsue* et non *corsu* (5.15). Quant à « vrai », l'occurrence *verai vertu* est suivie dans la même phrase d'une occurrence sans défaut de <e>, *veraie vertu* (5.13).

- (5.14) *douce pucele virge piue dame sanz venim et sanz fiel*
« douce jeune fille, vierge, dame pieuse sans venin ni fiel »
(mir, p. 34, trad. pers.)
- (5.15) *la raine fu corsue et grans*
« la reine était plantureuse et grande »
(fablesL, p. 4; fablesE1, p. 3; fablesT, p. 4; fablesY, p. 4, id.)⁶

Les autres lemmes sujets à variation peuvent se terminer par une ou deux consonnes. Lorsque les formes se terminent par deux consonnes, celles-ci sont toujours hétérosyllabiques, le groupe étant le plus souvent composé d'une liquide ou d'une nasale et d'une occlusive (5.16 a - c), parfois de /s/ et /t/ (5.16 d) ou de /r/ et /s/ (5.16 e). Ce type d'occurrences en défaut de <e> est illustré en (5.17) pour l'occurrence *rount*.

- (5.16) Groupes consonantiques précédant un défaut de <e>⁷
- (a) /rd/ *coarde ~ coart*
- (b) /nt/ *dolante ~ dolant*
gente ~ gent
reonde ~ rount
seinte ~ seint
quinte ~ quint
- (c) /ng/ *longe ~ long, lonc*
- (d) /st/ *preste ~ prest*
- (e) /rs/ *tierce ~ tierz*

jaco : PIERRE DE BEAUVAIS (1212). *Translacion mon seigneur saint Jaque*. Dees : Aube; compo. Beauvais, 1212; ms. champ., ca 1290; CR (BERKEY 1965); DEAF : PBeauvJacB.

5: Avec variantes graphiques vis-à-vis de fablesL.

fablesA : MARIE DE FRANCE (1180*). *Fables*. Ms. A; Dees : [Angleterre]; compo. NO, ca 1180; ms. a.n., ca 1250; MS (transcription inédite).

fablesB : MARIE DE FRANCE (1180*). *Fables*. Ms. B; Dees : [Angleterre]; compo. NO, ca 1180; ms. a.n., ca 1290; MS (transcription inédite).

fablesC : MARIE DE FRANCE (1180*). *Fables*. Ms. C; Dees : nil; compo. NO, ca 1180; ms. lorr., ca 1275; MS (transcription inédite).

calex : GUI (?) (1225). *La vie de sainte Catherine d'Alexandrie*. Ms. La Clayette; Dees : Région Parisienne; compo. pic., 1225; ms. nil, ca 1290; MS (TODD 1900); DEAF : SCathGuiT.

mir : GAUTIER DE COINCI (1224*). *Les miracles de Notre Dame de Soissons*. Ms. R; Dees : Aisne; compo. Soissonnais, ca 1224; ms. pic. mérid., ca 1310; CR (GAUTIER DE COINCI 1966); DEAF : CoincyI10K.

6: Avec variantes graphiques vis-à-vis de fablesL.

fablesL : MARIE DE FRANCE (1180*). *Fables*. Ms. L; Dees : nil; compo. NO, ca 1180; ms. nil, ca 1290; MS (transcription inédite).

fablesE1 : MARIE DE FRANCE (1180*). *Fables*. Ms. E; Dees : [Angleterre]; compo. NO, ca 1180; ms. a.n., ca 1225; MS (transcription inédite).

fablesT : MARIE DE FRANCE (1180*). *Fables*. Ms. T; Dees [Angleterre]; compo. NO, ca 1180; ms. champ., ca 1287; MS (transcription inédite).

fablesY : MARIE DE FRANCE (1180*). *Fables*. Ms. Y; Dees [Angleterre]; compo. NO, ca 1180; ms. norm. et a.n., ca 1225; MS (transcription inédite).

7: Il s'agit de *types*; les variantes graphiques non pertinentes ne sont pas représentées.

- (5.17) *el puiz d'enfern n ad porte mes vne fosse rount*
 « le puis infernal n'a pas de porte mais une fosse ronde »
 (sergb0, p. 3, trad. pers.)

Lorsque la seconde consonne du groupe, devenue finale suite à un défaut de <e>, est voisée, elle perd son voisement (comme au masculin) dans presque tous les cas : *coarde* devient *couart* (1 occ. / 1 occ. en défaut de <e>),⁸ *reonde* devient *rount* (1/1),⁹ en homographie avec les formes masculines. Seul le lemme « lonc » semble résister au dévoisement et/ou à la transcription graphique du dévoisement ; les 5 occurrences en défaut de <e> que nous avons conservées sont *lung* ou *long*.¹⁰ Pour ces lemmes, les formes paroxytones ont systématiquement la consonne sonore (7 formes *co(u)arde* ; 112 formes *reonde*, *roonde* ou *ronde* ; 316 formes *long(u)e*, *lung(u)e* ou *longhe*).

En laissant de côté la question du dévoisement final, ces groupes consonantiques sont tous attestés en position finale absolue en ancien français. Toutes catégories grammaticales confondues, les formes se terminant par ces groupes se comptent en milliers dans le NCA. Ainsi, 22 767 occurrences se terminent par <rt> (*mort*, *ert*, *part*, etc.), 198 812 occurrences se terminent par <nt> (*grant*, *quant*, etc., flexion verbale en *-nt* (*sont*) ou *-ent* (*furent*), adverbales en *-ment*, etc.), 1856 occurrences se terminent par <ng> (*loing*, *besoing*, *soing*, etc.), 75 294 occurrences se terminent en <st> (principalement dans la flexion verbale : *est*, *dist*, *fist*, etc., mais aussi *cest*, *tost*, *forest*, etc.) et 39 767 occurrences se terminent par <rs> (par ajout de *-s* casuel à une base en *-r* (*chevaliers*, *amors*) ou en *-rC* (*corp + s* → *cors*),¹¹ mais aussi *lors*, *fors*, etc.). Ces groupes sont tous largement attestés en ancien français en position finale, trait de voisement mis à part, et sont donc licites en cette position. En négatif, aucun défaut de <e> n'est constaté après un groupe *obstruante + liquide*, qui serait illicite en position finale, et ce même après réduction de celui-ci (on ne trouve pas, par exemple, de formes **autr* ou **aut* pour *autre*, **destr* ou **dest* pour *destre*, etc.).

5.2.2 Distribution diachronique générale

Les formes d'adjectifs féminins singuliers en défaut de <e> ne semblent pas suivre une dynamique diachronique très marquée, si ce n'est l'existence d'une légère concentration de textes disposant de taux moyens de défaut de <e> relativement élevés dans la seconde moitié du XII^e s, comme observable en fig. 5.1. C'est probablement cette concentration entre 1150 et 1200 qui génère la très légère pente descendante de la régression linéaire. Hormis ce léger effet, les autres détails de la distribution diachronique globale ne semblent pas significatifs ; la remontée de la régression locale à la fin de la période est principalement due à un faible nombre de textes après 1300, et au fait que les trois derniers textes ont des formes en défaut de <e>, trois textes ne pouvant être représentatifs d'une période à eux seuls.

Ce même effet de concentration autour de la seconde moitié du XII^e s. est à peine visible pour les occurrences en défaut de <e> situées devant initiale consonantique (fig. 5.10 a, *infra*), alors que l'effet est bien plus marqué pour les occurrences devant initiale vocalique (*id.* b). Néanmoins, devant initiale consonantique, l'effet redevient visible, quoique peu marqué,

sergb0 : GUISCHART DE BEAULIU (1190*). *Sermon*. Ms. O ; Dees : nil ; compo. a.n., ca 1190 ; ms. a.n., ca 1277 ; MS (GUISCHART DE BEAULIU 1909) ; DEAF : GuischartG.

8: contre, p. 52.

9: sergb0, p. 3.

10: chro, p. 28 ; edmund, p. 57 ; ipo, p. 14, 81, 117.

Nous n'avons pas conservé l'occurrence *lunc* de reis (p. 180), dans *e de lunc cele maisun*, où *lunc* est plus probablement la préposition pour « à côté de » (AND : « along(side); beside, near ; right up to »).

11: Pour la chute de C dans *rC+s* ; cf. SÉGÉRAL et SCHEER (2020, GGHF, p. 470).

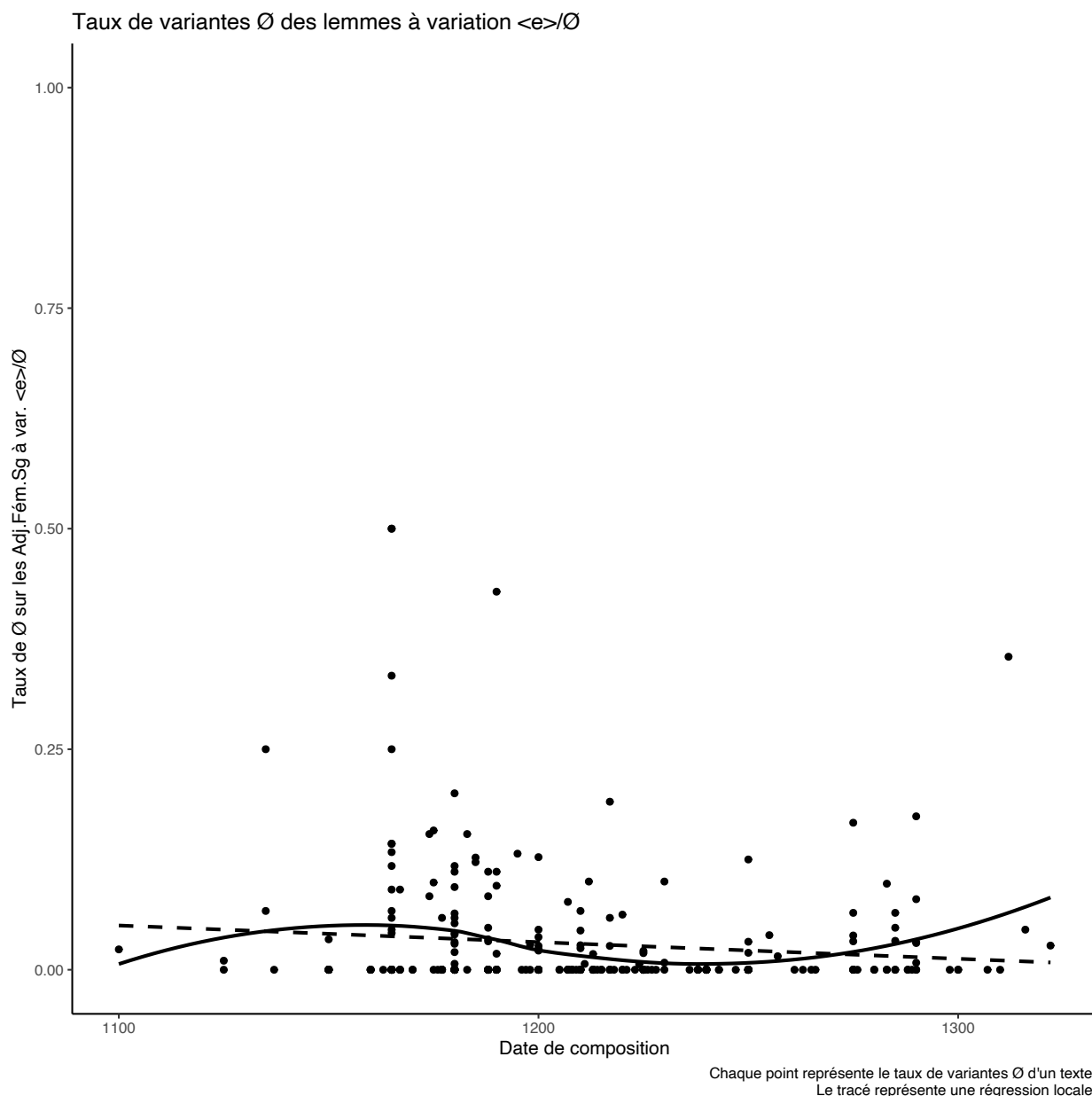
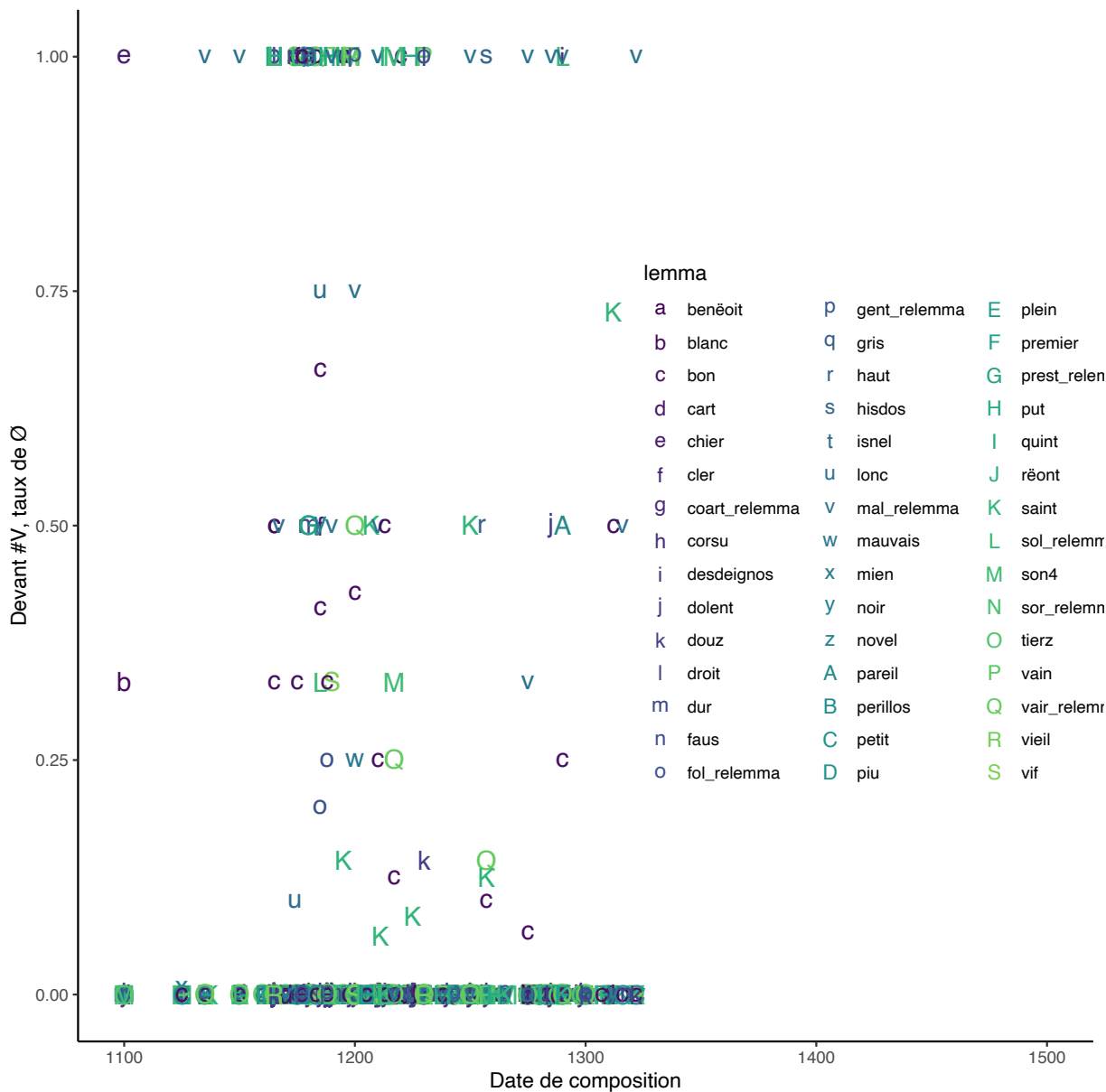


FIGURE 5.1 – Distribution contextuelle des variantes d'adjectifs féminins singuliers en défaut de <e>.

lorsque le tracé de la régression locale prend en compte la pondération des textes par leur nombre d'occurrences (*id. c*).

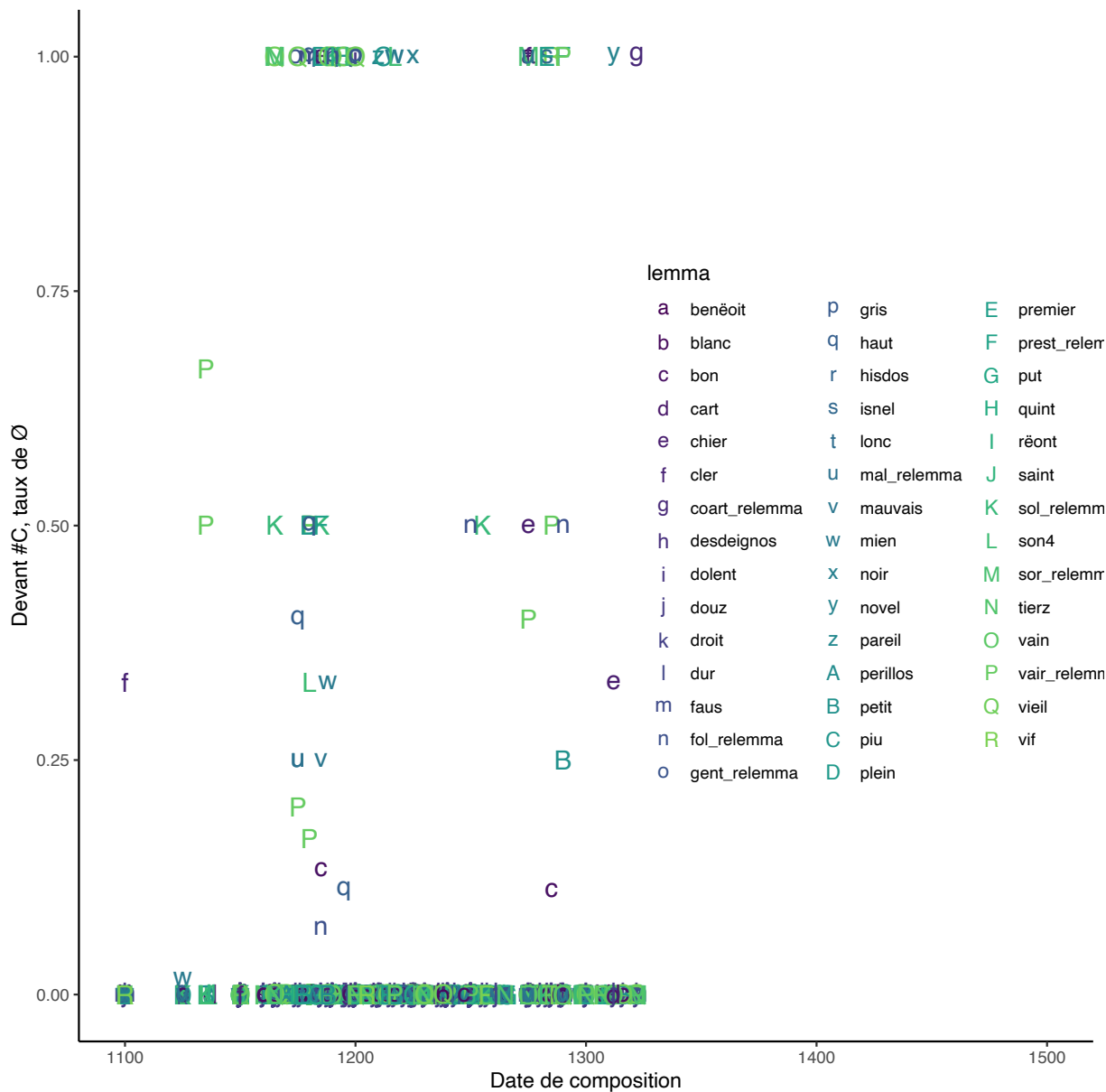
5.2.3 Distribution des lemmes en défaut de <e> : diachronie et fréquence

Les graphiques en fig. 5.2 et 5.3 représentent, pour chaque texte, le taux de défaut de <e> de chaque lemme, devant initiale vocalique en fig. 5.2, et devant initiale consonantique en fig. 5.3. On y observe que le premier texte du corpus, *roland*, ne dispose de défauts de <e> que pour trois lemmes : « cler », « blanc » et « chier » (fr.mod. « cher »). Le statut de



lemma	nb.e	nb.∅	tot.	tx.∅
desdeignos	1	1	2	0.50
corsu	4	3	7	0.43
pareil	2	1	3	0.33
droit	6	2	8	0.25
mal_relemma	96	26	122	0.21
put	23	6	29	0.21
quint	7	1	8	0.12
sol_relemma	55	7	62	0.11
son4	35	4	39	0.10
hisdos	10	1	11	0.09
[autres : 35]	2427	84	2511	0.03

FIGURE 5.2 – Distribution diachronique des alternances <e>/∅ selon les lemmes, adjectifs féminins singuliers suivis d’une initiale vocalique.



lemma	nb.e	nb.Ø	tot.	tx.Ø
gris	4	2	6	0.33
sor_relemma	10	2	12	0.17
coart_relemma	5	1	6	0.17
piu	5	1	6	0.17
isnel	7	1	8	0.12
prest_relemma	28	2	30	0.07
faus	61	3	64	0.05
pareil	23	1	24	0.04
vieil	213	9	222	0.04
son4	124	5	129	0.04
[autres : 34]	6000	65	6065	0.01

FIGURE 5.3 – Distribution diachronique des alternances <e>/<Ø> selon les lemmes, adjectifs féminins singuliers suivis d’une initiale consonantique.

l'occurrence en défaut de <e> du premier lemme est peut-être incertain, et nous permet d'illustrer un type de difficultés rencontrées pendant la phase d'exclusion des formes et lemmes. En effet, comme illustré en (5.18), l'adjectif apparaît déterminé mais sans nom, immédiatement après une première mention du syntagme *une clere buisine* auquel il renvoie.

- (5.18) *Met a sa buche une clere buisine,
Sunet la cler, que si païen l'oïrent ;*
« puis à sa bouche il met une claire trompette
« il la sonne clair et ses païens l'entendirent. »
ou : « il fait sonner la claire [trompette] »
(rolandox, p. 88, v. 3523-4, trad. SHORT (1990, p. 233) et pers.
après ou)

Dans cette position, et d'autant plus qu'il est précédé de *la* déterminant ou pronom, le mot *cler* peut être un adverbe (« clairement »), ce qui semble être le parti pris par SHORT (2007) dans sa traduction. Néanmoins, ce texte fait également partie du corpus de la BFM et il appartient au sous-corpus dont l'annotation a été vérifiée par des linguistes. Si la vérification humaine des plus de 29 000 mots du texte peut évidemment laisser passer quelques erreurs, le mot *cler* au v. 3524 est bien annoté ADJqua. Ainsi, rejeter cette occurrence reviendrait à supposer qu'une erreur d'annotation aurait échappé à Christiane Marchello-Nizia, responsable de la vérification des étiquettes sur ce texte. Aussi, nous préférons conserver cette occurrence.

La deuxième occurrence (5.19) est plus claire et dispose d'un statut intéressant : le syntagme *blanche ad la barbe* (« il a la barbe blanche ») revient deux autres fois dans le texte (5.20, 5.21), remplissant un hémistiche de quatre positions métriques. Dans ces deux autres occurrences du syntagme, l'adjectif n'est pas en défaut de <e>, alors que le <e> et le <h> qui le précède ont disparu au v. 3503.¹²

- (5.19) *blanc ad la barbe cume flur en avrill*
« sa barbe est blanche comme fleur en avril »
(rolandox, p. 88, v. 3503, trad. SHORT (1990, p. 231))
- (5.20) *blanche ad la barbe e tut flurit le chief*
« Sa barbe et blanche et son chef tout fleuri »
(*id.*, p. 3, v. 117, *id.*, p. 35)
- (5.21) *blanche ad la barbe ensemment cume flur*
« il a la barbe aussi blanche qu'une fleur »
(*id.*, p. 80, v. 3173, *id.*, p. 213)

Devant initiale vocalique, un défaut de <e> ne perturbe pas l'isosyllabisme du vers, aussi certains éditeurs considèrent que c'est l'élision qui est notée ici, la rendant par exemple par *Blanc'ad* (BFM rolandox : MOIGNET 1972b), tandis que d'autres peuvent être amenés à corriger le texte en rétablissant *blanche a*, comme le fait SHORT (1990) — mais il faut noter que l'édition de Short est une édition grand public, généralement dépourvue de notes d'édition des graphies. Le défaut de <h> ne doit pas, ici, être surinterprété, dans la mesure où la palatalité de la fricative n'est pas toujours indiquée dans le texte (*barbe blanche*, v. 3712).¹³

Cette pratique de l'ajout d'une apostrophe pour signifier une élision graphique exceptionnelle se retrouve dans la troisième occurrence de

12: La fricative ou affriquée notée <ch> est très rare en position finale absolue en a.fr. ; dans le NCA, <ch> final concerne essentiellement des noms propres étrangers (*Abimelech*, *Lamech*, etc.).

13: SHORT (2007, p. 116) indique que, dès *Roland*, <ch> et <c> peuvent être utilisés indifféremment pour noter /k/. Ce trait est mis en relation avec le traitement normanno-picard de *ca* initial ou interne postconsonantique, qui évolue en /ʃ/ au lieu de /k/ ; ainsi, *blanche* dans les autres dialectes donne, en picard, *blanke* (GOSSEN 1951, p. 74-75). Dans *Roland*, <c> et <ch> du FÉM *blanc(h)e* peuvent ainsi noter le traitement normanno-picard *ca* → /k/, à partir d'un étymon germanique †*blank* (FEW, TLFi), latinisé dès le x^e s. en *blancus*, -a (REY 2012).

défaut de <e> de rolandox. En (5.22), nous reproduisons l'édition du NCA, qui indique *cher amie*. L'édition de la BFM (MOIGNET 1972b) indique *cher'amie*, cette fois rejointe par SHORT (1990), qui indique *cher' amie* (avec une espace).

- (5.22) *soer cher amie de hume mort me demandes*
 « sœur, chère amie, tu parles là d'un homme mort »
 (rolandox, p. 93, v. 3713, trad. SHORT (1990, p. 243))

Ainsi, dès le premier texte de la période, trois lemmes sont concernés, deux devant initiale vocalique (le défaut de <e> est alors traité par les éditeurs comme une élision) et un devant initiale consonantique, mais dont l'occurrence est peu assurée. Face à ces occurrences en défaut de <e>, Comparalem a détecté dans le même texte six occurrences paroxytones *clere*,¹⁴ 11 occurrences paroxytones *blanc(h)e*¹⁵ et 1 occurrence *chere*.¹⁶

Les deux lemmes suivants à connaître le défaut de <e> sont « mien » et « vair/verai ». L'adjectif possessif « mien » apparaît en défaut de <e> à trois reprises dans psautier (5.23 - 5.25), texte anglo-normand de la première moitié du XII^e s.

- (5.23) *a la mei larme ne te assurdisses*
 « ne te fais pas sourd à ma larme »
 Litt. « à la larme mienne [...] »
 (psautier, p. 32, trad. pers.)
- (5.24) *a encurvoer la mei aneme*
 « à infléchir mon âme »
 Litt. « [...] l'âme mienne »
 (*id.*, p. 46, *id.*)
- (5.25) *la meie force e la mei loenge*
 « ma force et ma louenge »
 Litt. « la force mienne et la louenge mienne »
 (*id.*, p. 101, *id.*)

Ces occurrences ne laissent guère de place à l'interprétation du mot comme autre chose qu'un adjectif possessif féminin singulier. Morphologiquement parlant, la diphtongaison de ces occurrences indique qu'il ne peut s'agir de déterminants possessifs (MARCELLO-NIZIA 2020b, GGHF, p. 702-704); syntaxiquement, l'absence de double détermination (*la mei la larme*, « la larme, la mienne ») indique que *mei* appartient au syntagme déterminé *la + nom*, le pronom possessif devant être déterminé (MARCELLO-NIZIA 2020b, GGHF, p. 711). Face à ces trois occurrences en défaut de <e>, psautier dispose de 246 occurrences *meie* et d'une occurrence *maeie*; son point (en forme de 'x' ou de 'w') ne se détache donc qu'à peine des lemmes à 0% de défaut en fig. 5.2 et 5.3, comme mis en valeur en figures marginales 5.4 et 5.5. Par ailleurs, les syntagmes correspondant aux trois occurrences en défaut de <e> apparaissent sans défaut de <e> dans oxfps (5.26 - 5.28), un psautier anglo-normand contemporain présent dans la BFM; certes au pluriel en (5.26).

- (5.26) *ot tes oreiles receif les meies lermes*
 « que tes oreilles perçoivent mes larmes »
 (oxfps, p. 22, xxxviii, l. 16, trad. pers.)

14: V. 59, 445, 737, 2316, 2512, 3523.

15: V. 117, 1494, 1843, 2316, 2334, 2930, 2943, 3173, 3520, 3618, 4001.

16: V. 3031.

psautier : ANON. (1125*). *Livre des psaumes*. Dees nil; compo. a.n., ca 1125; ms. a.n., ca 1150; CR2 (MICHEL 1876), DEAF: PsCambrM

MARCELLO-NIZIA (2020). « Les possessifs — déterminants, adjectifs et pronoms ».

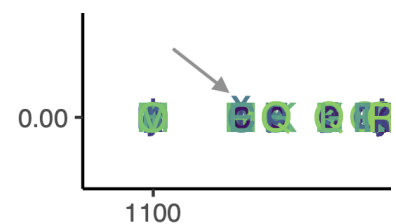


FIGURE 5.4 – Extrait de la fig. 5.2, avec mise en valeur du point (en forme de 'x' du lemme « mien » dans psautier.)

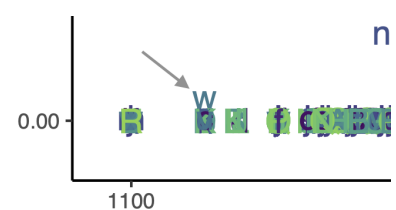


FIGURE 5.5 – Extrait de la fig. 5.3, avec mise en valeur du point (en forme de 'w' du lemme « mien » dans psautier.)

oxfps : ANON. (ca 1125*). *Psautier d'Oxford*. Compo. 1^{ère} m. XII^e s., ms. m. XII^e s., dial. a.n. (MICHEL 1860).

- (5.27) *e encuroerent la meie aneme.*
« et [ils] infléchirent mon âme. »
(*id.*, p. 32, lvi, l. 8, *id.*)
- (5.28) *La meie fortece et la meie loenge*
« ma force et ma louenge »
Litt. « la force mienne et la louenge mienne »
(*id.*, p. 68, cxvii, l. 14, *id.*)

Le lemme *vair_*relemma, issu de la fusion des étiquettes *vair1*, *voir* et *verai*, correspond à l'adjectif « vrai, véritable » en fr.mod. Ce lemme apparaît en défaut de <e> dans *marga* et *mrgr1* (5.29 - 5.31). Ces deux textes représentent des versions contemporaines de la *Vie de sainte Marguerite* de Wace, légèrement postérieures à *psautier*, et localisées dans la Haute-Marne par Dees¹⁷ mais de composition normande selon la bibliographie philologique.

- (5.29) *verai salvere vrai jugieres vrai justice*
« Vrai sauveur, vrai juge, vraie justice »
(*marga*, p. 10, trad. pers.)
- (5.30) *voirs sauviere voire esperance voirs pitez*
« vrai sauveur, vraie espérance, vraie pitié »
(*mrgr1*, p. 5, *id.*)
- (5.31) *voirs secors voirs jugiere voirs justise*
« vrai secours, vrai juge, vraie justice »
(*id.*, p. 5, *id.*)

Les deux occurrences de *mrgr1* se trouvent dans la même longue énumération de qualificatifs se rapportant à Dieu, qui correspondent également à l'occurrence de *marga*. Ainsi, on observe ici un défaut de <e> dans le même syntagme (*verai/voirs justice*), auquel s'ajoute un second défaut dans *mrgr1* (*voirs pitez*). Le statut des formes est néanmoins différent : dans *marga*, il s'agit d'un simple défaut de <e> (*veraie* tronqué en *verai*), tandis que dans *mrgr1* la forme défectueuse semble alignée sur le masculin (*voire* tronqué en *voir*; ajout du -s flexionnel du masc. CSsg). Face à ces trois occurrences du lemme « vrai/vair », *marga* dispose d'une occurrence *vraie* et d'une occurrence *voire*,¹⁸ et *mrgr1* dispose de deux occurrences *voire*.¹⁹

À partir du milieu du xii^e s., de nombreux lemmes deviennent attestés en défaut de <e>. 25 nouveaux lemmes sont attestés en défaut de <e> avant 1200 (5.32), tandis que quatre lemmes déjà atteints ont encore des occurrences en défaut de <e> (5.33). En revanche, le lemme « blanc » ne reparait pas en défaut de <e>, l'occurrence de *roland* en (5.19) étant la seule occurrence défectueuse du corpus. Plus de la moitié des lemmes est donc attestée en défaut de <e> dans le premier siècle de notre corpus.

marga : WACE (1135*). *La vie de Sainte Marguerite de Wace*. Ms. M; Dees : Haute-Marne; compo. norm., ca 1135; ms. art., 1267; CR3 (WACE 1932a); DEAF : [WaceMargaF](#).

mrgr1 : WACE (1135*). *Vie de Sainte Marguerite*. Ms. M; Dees : [Haute-Marne]; compo. norm., ca 1135; ms. tour., ca 1238; CR3 (WACE 1932b); DEAF : [WaceMarmF](#).

17: Pour *mrgr1*, l'indication *RegionDees* et le coefficient dialectal sont manquants, mais le *codeRegional* présent indique la Haute-Marne.

18: p. 10, p. 5.

19: p. 1, p. 5 (5.30).

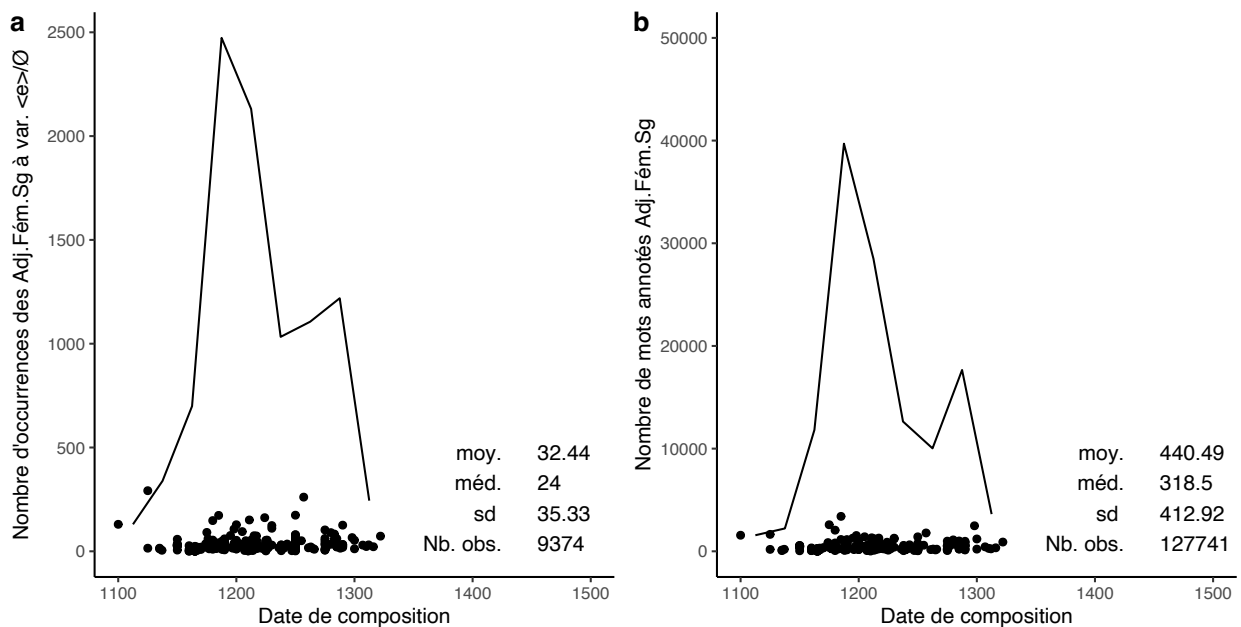
- (5.32) Lemmes atteints dans la seconde moitié du XII^e s. :
 « bon », « cart »,²⁰ « cler », « corsu », « droit », « dur », « faus »,
 « fol » (fol_relemma, cf. (5.10)), « gent » (gent_relemma),
 « haut », « lonc », « mal » (mal_relemma), « noir », « perillos »,
 « petit », « plein », « prest » (prest_relemma), « put »,²¹ « rëont »,
 « saint », « sol » (sol_relemma),²² « son », « tierz », « viel », « vif ».
- (5.33) Lemmes atteints préalablement, et en défaut de nouveau :
 « chier », « cler », « mien », « verai/vair ».

20: *cart* : fr.mod. « quatrième ».

21: *put* : fr.mod. « corrompu, décadent »
 (AND : « loathsome, vile, decaying ; despicable, base, morally corrupt ; whorish, lascivious »).

22: *sol* : fr.mod. *seul*.

Il est peu probable que cette croissance soudaine du nombre de lemmes attestés en défaut de <e> soit à analyser comme une brusque accélération de la diffusion du défaut de <e> au sein du lexique ; elle tient plus probablement à la diachronie du corpus lui-même, avec une explosion du nombre de textes dans la seconde moitié du XII^e s. Ainsi, comme le montrent les deux graphiques en fig. 5.6, l'évolution du nombre d'occurrences des 45 lemmes sélectionnés (a) est globalement parallèle à celle du nombre d'adjectifs féminins singuliers dans le corpus (b), avec une brusque croissance dans la seconde moitié du XII^e s. Sur ces graphiques, chaque point représente le nombre d'occurrences d'un texte tandis que la courbe représente le cumul d'occurrences sur une fenêtre incrémentale de 25 ans. Il ne semble donc pas y avoir, ici, de diffusion progressive du défaut de <e> dans le corps lexical, mais plutôt une attestation d'un nombre croissant de lemmes, qui permet aux formes en défaut de <e> d'apparaître.



Chaque point représente un texte.

Les courbes relient la somme des occurrences sur une fenêtre de 25 ans.

FIGURE 5.6 – Distribution diachronique des occurrences au féminin singulier des lemmes adjectivaux à défaut de <e> (a), en regard de distribution diachronique des occurrences de tous les lemmes d'adjectifs féminins singuliers (b).

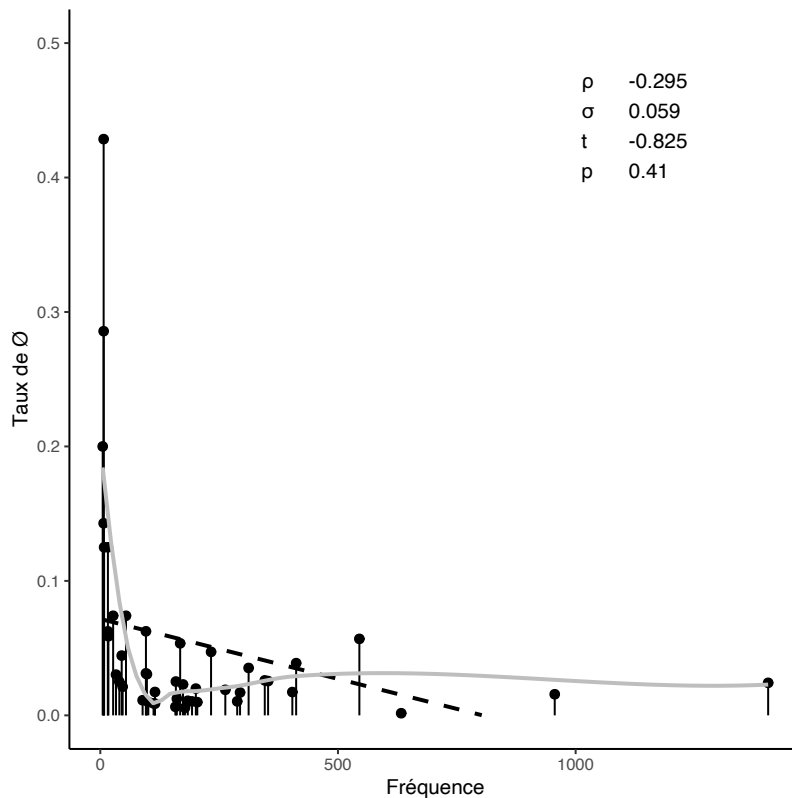


FIGURE 5.7 – Distribution des lemmes d’adjectifs féminins singuliers en fonction de leur fréquence et de leurs taux de défaut de <e>.

5.2.3.1 Fréquence des lemmes et défaut de <e>

Par ailleurs, la fréquence des lemmes sujets à défaut de <e> ne semble pas avoir un effet robuste sur le taux de défaut de leurs occurrences. Selon les études consacrées à l’effet de la fréquence lexicale, plus un lemme est fréquent, plus il est sujet à la chute de certains segments ; c’est le REDUCING EFFECT, lié à une vitesse de traitement mental proportionnelle à la fréquence des items (BYBEE 2006, p. 714-715, 2000, 2002 ; BYBEE et SCHEIBMAN 1999). Ici, cet effet ne semble pas représenté, et, en première lecture, c’est même l’inverse dont semblent témoigner nos données.

En fig. 5.7, chaque lemme est représenté par un point, dont la position sur l’axe des abscisses correspond à la fréquence, tandis que la position sur l’axe des ordonnées correspond au taux de défaut de <e> de ce lemme ; les données qui génèrent cette figure sont reprises en table 5.2, page suivante. La régression linéaire sur ce graphique indique une tendance globalement décroissante : le taux de défaut de <e> est plus ou moins proportionnellement inverse à la fréquence des lemmes. Mais cette corrélation est plutôt faible ($\rho = -0,30$) et non significative ($p = 0,41$). Néanmoins, on peut observer qu’au moins la partie gauche de la courbe de régression locale semble adopter un profil de décroissance exponentielle, semblable à celui de l’arc d’hyperbole en figure marginale 5.8.

Cette ressemblance est d’autant plus marquante si l’on observe la série

BYBEE (2006). « From Usage to Grammar: The Mind’s Response to Repetition ».
 — (2000). « The phonology of the lexicon: Evidence from lexical diffusion ».
 — (2002). « Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change ».
 BYBEE et SCHEIBMAN (1999). « The effect of usage on degrees of constituency: the Reduction of *don’t* in English ».

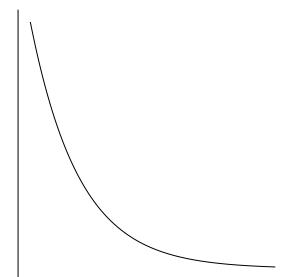


FIGURE 5.8 – Arc d’hyperbole de fonction $y = 1/x$.

TABLE 5.2 – Relation entre fréquence et taux de défaut de <e>, pour chaque lemme d'adjectif féminin singulier.

lemma	nb.Ø	nb.E	frq.	tx.Ø	lemma	nb.Ø	nb.E	frq.	tx.Ø
bon	34	1371	1405	2%	tierz	1	157	158	1%
saint	14	941	955	1%	cart	1	114	115	1%
douz	1	632	633	0%	noir	2	113	115	2%
mal_relemma	31	514	545	6%	vain	1	111	112	1%
mien	7	397	404	2%	novel	1	100	101	1%
fol_relemma	9	344	353	3%	vif	3	95	98	3%
haut	9	337	346	3%	put	6	90	96	6%
vieil	11	301	312	4%	faus	3	93	96	3%
lonc	5	289	294	1%	rëont	1	88	89	1%
premier	3	285	288	1%	prest_relemma	4	50	54	7%
cler	5	258	263	2%	benëoit	1	46	47	2%
sol_relemma	11	222	233	5%	droit	2	43	45	4%
vair_relemma	9	218	227	4%	perillos	1	39	40	3%
mauvais	2	202	204	1%	quint	1	32	33	3%
chier	4	197	201	2%	pareil	2	25	27	7%
petit	2	191	193	1%	isnel	1	16	17	6%
dolent	2	182	184	1%	hisdos	1	15	16	6%
verai	5	178	183	3%	sor_relemma	2	14	16	13%
blanc	1	177	178	1%	coart_relemma	1	7	8	13%
gent_relemma	4	170	174	2%	corsu	3	4	7	43%
son4	9	159	168	5%	gris	2	5	7	29%
dur	2	159	161	1%	piu	1	6	7	14%
plein	4	155	159	3%	desdeignos	1	4	5	20%

de points grisés en figure marginale 5.9, qui suit d'assez près le profil de décroissance exponentielle. Plutôt que d'indiquer un phénomène linguistique lié à l'effet de la fréquence sur le traitement grapho-phonologique des itérations des items lexicaux, ce profil semble ici indiquer un effet de filtre aléatoire : moins un lemme est fréquent, plus la probabilité d'un taux très élevé de défaut de <e> augmente, simplement parce qu'un lemme peu fréquent peut obtenir un taux très élevé avec seulement quelques occurrences en défaut de <e>. Ainsi, pour « corsu » (3 occurrences en défaut sur 7 occurrences : 43%) et « gris » (2 occurrences en défaut sur 7 occurrences : 29%), la corrélation *faible fréquence* \Leftrightarrow *fort taux de défaut de <e>* ne tient probablement qu'au poids plus important du hasard dans l'établissement du taux des lemmes à faible fréquence. Il y a probablement bien d'autres lemmes à faible fréquence qui seraient susceptibles de comporter des défauts de <e> si leur fréquence avait été plus élevée, mais qui ne sont de fait pas attestés dans nos données. Au contraire, plus un lemme dispose d'une fréquence élevée, moins l'effet de filtre est probable ; tous nos lemmes disposant de plus de 100 occurrences (28 lemmes) ont des taux de défaut de <e> compris entre 0,16% et 5%, loin du maximum de 43% atteint à 7 (« corsu »).

Ainsi, plus de la moitié des lemmes est attestée dès lors que le corpus devient conséquent, ce qui ne plaide pas en faveur d'une analyse précise de la diffusion lexicale des défauts de <e> : ceux-ci semblent déjà présents dans la langue du corpus, et n'attendent qu'un nombre suffisant d'occurrences pour apparaître. La corrélation entre la fréquence des

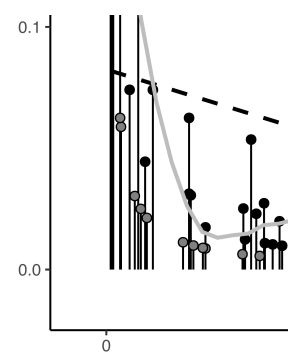


FIGURE 5.9 – Extrait de la fig. 5.7, avec mise en valeur de la disposition hyperbolique de certains points.

lemmes et leur taux de défaut de <e> est non significative, et son dessin hyperbolique inverse témoigne probablement d'un effet de filtre : moins un lemme est fréquent, plus l'apparition de quelques rares formes en défaut de <e> est à même de générer des taux de défaut de <e> très élevés.

5.2.3.2 Fréquence, collocations et défauts de <e>

Un troisième effet de fréquence peut être observé dans nos données : celui de la collocation. En effet, pour trois lemmes (« saint, lonc, mal »), les défauts de <e> apparaissent fréquemment dans des constructions récurrentes. La collocation est en principe associée à un statut particulier dans le traitement mental des items linguistiques, lié à une réduction phonologique plus marquée que dans la concaténation d'items lexicaux indépendants (BYBEE 2006, p. 723-725), à cause d'une part du REDUCING EFFECT de la simple fréquence, mais aussi du traitement global des items de la construction, pouvant résulter en un affaiblissement, voire en une abolition des frontières de mots au sein du syntagme (BYBEE et SCHEIBMAN 1999; BYBEE 2006) — en cas d'abolition totale des frontières de mots, lorsque la collocation constitue une seule entrée lexicale, on parle alors de *token*. Nous avons repéré la récurrence de certains syntagmes *adjectif + nom* contenant des défauts de <e> ; dans ce qui suit nous les abordons lemme par lemme.

Saint(e) + nom. Ainsi, sur les 14 occurrences en défaut de <e> du lemme « saint » retenues dans les résultats de Comparalem, 12 correspondent au syntagme *sainte Église*, tel qu'illustré en (5.34) et (5.35). S'y ajoutent trois occurrences *seint iglise* (5.36), exclues de nos résultats puisque le graphème initial <i> peut, selon les mots, représenter une voyelle ou une consonne (ce qui n'est évidemment pas le cas ici, mais est une restriction générale de nos requêtes, cf. sec. 3.1.1). L'adjectif féminin singulier « saint » n'apparaît en défaut de <e> hors du syntagme *sainte Église* que dans deux occurrences, également devant initiale vocalique (5.37 - 5.38).

- (5.34) *il suffrit le sacrefice de sey pur dieu e seint eglise*
« il endura d'être sacrifié, pour Dieu et la sainte Église »
(edmund, p. 48, trad. pers.)
- (5.35) *s en est alee a seint yglise*
« [la dame] est allée à l'église sainte »
(fabd, p. 2, trad. pers.)
- (5.36) *anorer toz iorz seint iglise*
« honorer toujours la sainte Église »
(ailed, p. 6, trad. pers.)
- (5.37) *la feste de seint agnes*
« la fête de saint Agnès »
(ndchar, p. 92, trad. pers.)
- (5.38) *il desputent e jours e nuis de seint escripture*
« ils se querellent jour et nuit à propos des saintes écritures »
(plainte, p. 6, trad. pers.)

edmund : DENIS PIRAMUS (1195*). *La Vie Seint Edmund le Rei*. Dees : [Angleterre]; compo. a.n., ca 1195; ms. a.n., ca 1325; MS2 (DENIS PIRAMUS 1935); DEAF : EdmK.

fabd : ANON. (1250*). *Fabliaux n°2*, 4. Ms. D; Dees : Haute-Marne; compo. pic., ca 1250; ms. frc., ca 1290; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).

ailed : RAOUL DE HOUDENC (1207*). *Le roman des ailes*. Ms. D; Dees : Aube; compo. pic., ca 1207; ms. frc., ca 1290; MS (RAOUL DE HOUDENC 1983).

ndchar : JEAN LE MARCHANZ (1157*). *Miracles de Notre Dame de Chartres*. Dees : Orléanais; compo. Chartres, ca 1157; ms. Chartres, ca 1350; MS2 (JEAN LE MARCHANZ 1973); DEAF : MirNDChartrK.

plainte : NICHOLAS BOZON (1312). *La plainte d'amour*. Dees : nil; compo. a.n., 1312; ms. a.n., ca 1325; CR (BOZON 1905); DEAF : PlainteAmV.

Dans les deux derniers exemples, si l'adjectif n'appartient pas à la collocation *sainte Église*, il n'appartient pas à n'importe quel syntagme

non plus. En (5.37), *seint agnes* représente la structure *saint + nom de saint*, qui est fort régulière; c'est par ailleurs dans cette structure que SÉGÉRAL et SCHEER (2020, GGHF, p. 407) repèrent une chute de *-t* dans l'occurrence *sein Gabriel (Roland, v. 2847)*, analysée comme un effet de sandhi externe. Enfin, dans le dernier exemple, si le syntagme *seint escripture* n'est présent en défaut de <e> que dans une occurrence dans notre corpus, il appartient également à une collocation, le mot « écriture » étant précédé de l'adjectif *saint* avec une certaine récurrence, pour désigner la Bible.²³

Long(e) + nom. Le lemme « lonc » semble également être affecté par la collocation dans notre corpus. Ainsi, sur les 5 occurrences en défaut de <e> du lemme, trois sont suivies de *aleine*, dans l'équivalent du fr.mod. [*de*] *longue haleine*, avec un sens encore rattaché concrètement au souffle (5.39 - 5.41).²⁴

- (5.39) *a long aleine bien soneie*
« il sonnait bien, d'un souffle long [sans reprendre son souffle] »
(ipo, p. 14, trad. pers.)
- (5.40) *un lung aleine un corn suna*
« il fit sonner un cor d'un souffle long [id.] »
(ipo, p. 81, trad. pers.)
- (5.41) *od lung aleine un corn suna*
« il fit sonner un cor d'une longue haleine [id.] »
(ipo, p. 117, trad. pers.)

Les deux autres occurrences du lemme en défaut de <e> (5.42, 5.43), en revanche, ne semblent pas appartenir à une collocation particulière.²⁵ Une nouvelle fois, ce fonctionnement de collocation ne concerne donc pas toutes les occurrences en défaut de <e> du lemme, et le défaut de <e> ne concerne pas non plus toutes les occurrences de la collocation : sur les 11 occurrences de la collocation *longue aleine*, 8 disposent bien d'un <e> (*long(u)e aleine*). Le texte qui contient les trois occurrences en défaut de <e> de la collocation, ipo, contient lui-même une occurrence où le <e> ne fait pas défaut (5.44), ce qui signifie que même pour le scribe de ce texte, la collocation n'implique pas nécessairement le défaut de <e>.

- (5.42) *long espee*
« longue épée »
(chro, p. 28, trad. pers.)
- (5.43) *il vnt l eglise apresté mult bele e grant e long e le e auenant*
« ils ont préparée l'église, très belle, grande, longue, large et accueillante »
(edmund, p. 56-57, trad. pers.)
- (5.44) *a buche met un olifan a lunge aleine l at suné*
« il porte à sa bouche un olifant; d'un souffle long il l'a fait sonner »
(ipo, p. 148, trad. pers.)

mal + nom. Enfin, le lemme « mal » semble également répondre à un effet de la collocation pour les défauts de <e>. Sur les 25 occurrences en défaut de <e> que nous avons conservées, 12 correspondent au syntagme *mal heure*, ancêtre du fr.mod. *malheur*, avec encore le sens d'un moment

SÉGÉRAL et al. (2020). « Phonétique Historique ».

23: Dans le NCA, le lemme « écriture » est précédé du lemme « saint » dans 49 occurrences, sur 282 occurrences du lemme.

24: DMF « De tout son souffle; sans reprendre son souffle »; TLFi *lunge aleine* pour le souffle dans le cor dans *Roland*. Dans le NCA, le lemme « aleine » est précédé du lemme « lonc » dans 11 occurrences sur 66.

ipo : HUE DE ROTELANDE (ca 1185). *Ipomedon*. Compo. a.n., ca 1185; ms. a.n., ca 1190; CR (transcription inédite).

25: Dans le NCA, le lemme « espee » est précédé du lemme « lonc » dans seulement 17 occurrences sur 1158, tandis que l'énumération en (5.43) n'est guère dotée d'une fréquence satisfaisante, même en considérant uniquement « long » suivi de « et » (169 occurrences sur 1261 occ. du lemme « lonc »).

chro : BENOIT DE SAINTE-MAURE (1174*). *Chronique des ducs de Normandie*. Ms. [Tours, Bibl. Munic. 903]; Dees : [Vendée, Deux-Sèvres]; compo. poit., ca 1174; ms. a.n., ca 1217; CR2 (BENOIT DE SAINTE-MAURE 1951-1967); DEAF : BenDucF.

edmund : DENIS PIRAMUS (1195*). *La Vie Saint Edmund le Rei*. Dees : [Angleterre]; compo. a.n., ca 1195; ms. a.n., ca 1325; MS2 (DENIS PIRAMUS 1935); DEAF : EdmK.

inapproprié ou d'une circonstance malheureuse, souvent associée à la naissance (5.45 - 5.48), mais aussi parfois à des déplacements, physiques (5.49) ou abstraits (5.50); dans ce dernier exemple, le déplacement correspond au va et vient des individus en termes de richesse et de vertu.

- (5.45) *comme a mal heure vous nasquistes*
« puisque vous êtes nés à un mauvais moment [= sous de mauvais auspices] »
(bar, p. 11, trad. pers.)
- (5.46) *a mal ure fu nez*
« il naquit à un mauvais moment [= sous de mauvais auspices] »
(sergbH, p. 4, trad. pers.)
- (5.47) *a com mal eure vous venistes onques en tere*
« comme vous êtes venu sur terre à un mauvais moment [= sous de mauvais auspices] »
(robert, p. 15, trad. pers.)
- (5.48) *com mar fui nés et com de mal eure engentrés*
« puisque je suis né sous de mauvais auspices et que je fus engendré à un mauvais moment [= sous de mauvais auspices] »
(amad, p. 22, trad. pers.)
- (5.49) *car de mal eure i fist entree*
« car il arriva à un mauvais moment [= à un moment inapproprié] »
(gepa, p. 8, trad. pers.)
- (5.50) *cum il plus muntent plus descendent [...] m est vis cil a mal ure munte*
« puisque plus ils montent, plus ils descendent [...] il m'est avis que celui-ci monte à un mauvais moment [= sous de mauvais auspices] »
(ipo, p. 148, trad. pers.)

bar : ANON. (1210*). *Le chevalier au barisel*. Dees : Somme, Pas-de-Calais ; compo. pic., ca 1210 ; ms. pic., 1310 ; CR2 (ANON. 1955) ; DEAF : ChevBarAnL.

sergbH : GUICHART DE BEAULIU (ca 1190). *Sermon*. Compo. a.n., ca 1190 ; ms. a.n., ca 1250 ; CR1 (BEAULIU 1909), DEAF : GuischartG.

robert : ANON. (1190*). *Robert le diable*. Dees : Somme, Pas-de-Calais ; compo. traits nomr.?, ca 1190 ; ms. pic., ca 1275 ; CR (ANON. 1903) ; DEAF : RobDiablL.

amad : ANON. (1210*). *Amadas et Ydoine*. Ms. P ; Dees : Aisne ; compo. a.n., ca 1210 ; ms. pic., ca 1289 ; CR3 (ANON. 1974) ; DEAF : AmYdR.

gepa : GEOFFROY DE PARIS (1316). *Chronique métrique*. Dees : Normandie ; compo. norm., 1316 ; ms. nil, ca 1318 ; MS (GEOFFROY DE PARIS 1956) ; DEAF : GeoffrParChronD.

Encore une fois, la seule collocation *mal heure* ne saurait être responsable de tous les défauts de <e> ; dans 13 occurrences, *mal* n'est pas suivi de « heure ». Si certaines occurrences semblent appartenir à d'autres collocations, il n'est pas certain que toutes les occurrences se rapportent à une collocation en tant que telle, ou quel est le degré de cohésion entre les items impliqués. Ainsi, dans deux occurrences, *mal* précède *estrine* (DMF : « présage, augure, sort (favorable ou funeste) »), la collocation *mal estrenne* prenant le sens de « faire qqc. sous de mauvais auspices, pour son malheur/être dans une mauvaise situation » (DMF) (5.51), parfois dans une formule de malédiction (5.52). De même, *mal aventure* en (5.53) correspond visiblement à un usage lexicalisé, attesté parfois même sous la forme d'un seul mot (AND, DMF « malaventure » ; AND dès 1150, « calamity, misfortune; misery, unhappiness; wickedness »).

- (5.51) *tous li mons doit hui juner et vous vous volés desjuner et char mengier a mal estrine*
« tout le monde doit jeûner aujourd’hui et vous, voulez rompre le jeûne et manger de la viande pour votre malheur »
(bar, p. 11, trad. pers.)
- (5.52) *diex vous doit mal estrine quant si desordenement parlés*
« Dieu vous donne un mauvais présage quand vous parlez avec tant de confusion »²⁶
(gar, p. 3, trad. pers.)
- (5.53) *lo fol croit si l en uient domaige es laz chiet de mal aventure*
« le fou croit, lorsqu’il est en péril, qu’il est tombé dans le piège d’une mauvaise aventure »
(yzop, p. 22, trad. pers.)

Dans d’autres cas, la construction *mal + nom* se rapproche de locutions enregistrées par les dictionnaires, mais sans que la correspondance soit exacte. Ainsi, en (5.54), le syntagme *mal entente* peut être rapproché de *venir a male entente*,²⁷ des formes lexicalisées *mesentendant*²⁸ et *malentendu*,²⁹ d’équivalents sémantiques comme *male fin* ou *male volonté*,³⁰ et opposable à *en/par bonne entente*,³¹ mais *mal entente* n’est pas attesté tel quel par les dictionnaires que nous avons consultés (DMF, AND, GD, FEW).

- (5.54) *jo ai grant pitie quant crestien sunt mesaisié quant crestien unt mal entente*
« j’ai beaucoup de compassion quand des chrétiens sont en détresse parce qu’ils ont de mauvaises intentions »
(teo2, p. 14, trad. pers.)

Ici, la très forte productivité de *mal + nom* dans le domaine de la volonté, du dessein, de l’intention, brouille visiblement les pistes entre ce qui tient de possibilités combinatoires étendues et ce qui tient de la collocation à proprement parler, voire de la lexicalisation. Ceci est d’autant plus problématique que l’effet de la fréquence des *tokens* (ici par le biais de la formation des collocations) et celui de la fréquence des *types* s’opposent : plus un *type* peut être construit avec des items différents, moins il fonctionne comme un *token* (BYBEE et SCHEIBMAN 1999, p. 578; DIVJAK 2019, p. 45). Les potentialités de collocation de *mal + nom* ne se limitent d’ailleurs pas au temporel et au dessein ; de fait, dans les usages strictement adjectivaux, le GD documente pas moins de 17 locutions en *mal + nom*.³² On retrouve une situation probablement similaire pour *en mal envie* en (5.55) et *de mal uxor* en (5.56), où la frontière entre productivité des associations et formation de collocations se brouille.

- (5.55) *k itant est fel et envius vit cestui tant religios mult li pesa de sa bunte en mal envie l ad mué el queor*
« [celui] qui est si traître et jaloux vit l’autre être si religieux [qu’il souffrit beaucoup de sa bonté et la tranforme en mauvais désir en son cœur »
(gepa, p. 8, trad. pers.)
- (5.56) *jo ai guerpi mun criator par le conseil de mal uxor*
« j’ai déserté mon créateur sur le conseil d’une mauvaise femme »
(myst, p. 7, trad. pers.)

26: Cf. DMF « étrenne » A.2.c : « [Dans une formule de malédiction] [...] *Que Dieu lui doint hui mal estrainne!* [...] *Dieu leur envoie mal estrainne!* »

gar : ANON. (1275*). *Le Garçon et l’aveugle*. Dees : Aisne ; compo. pic., ca 1275 ; ms. pic., ca 1275 ; CR2 (ANON. 1969 [1921]) ; DEAF : *GarçAvR*².

yzop : ANON. (1275*). *Ysopet de Lyon*. Dees : Franche-Comté ; compo. Est, ca 1275 ; ms. frcomt., ca 1290 ; MS (ANON. 1882) ; DEAF : *YsLyonF*.

27: AND : « to achieve one’s evil intentions ».

28: FEW « întëndère » 3.a §2 : « mal intentionné, pervers ».

29: FEW *id.* §3 : « mal intentionné ».

30: GD *male fin* : « mauvaise fin » ; *male volonté* : « mauvais dessein, conspiration »

31: DMF : « dans de bonnes intentions, dans de bonnes dispositions »

teo2 : ADGAR (1150*). *Le dit de Théophile*. Ms. M ; Dees : nil ; compo. a.n., ca 1150 ; ms. a.n., ca 1210 ; MS2 (WEBER 1877).

DIVJAK (2019). *Frequency in Language: Memory, Attention and Learning*

32: *male action* (« terme de droit »), *male amour* (« antipathie, mauvais sentiments »), *mal an* (« mauvaise année ») et *mettre en mal an* (« accabler de maux »), *mal art* (« artifice, tromperie »), *male façon* (« s’est souvent employé comme substantif composé pour signifier méfait, mauvaise action »), *male fin* (« mauvaise fin »), *male foi* (« mauvaise foi »), *male gote* (« sorte de maladie »), *male grace* (« disgrâce »), *male hart* (« corde à pendre »), *male honte* (« sorte de jeu »), *mau jour* (« mauvais jour, malheur »), *male paix* (« mécontentement »), *mal point* (« mauvaise situation »), *male rage* (« rafe furieuse »), *male volonté* (« mauvais dessin, conspiration ») et *male semaine* (« mensuration »).

gepa : GEOFFROY DE PARIS (1316). *Chronique métrique*. Dees : Normandie ; compo. norm., 1316 ; ms. nil, ca 1318 ; MS (GEOFFROY DE PARIS 1956) ; DEAF : *GeoffrParChronD*.

myst : ANON. (1175*). *Le mystère d’Adam*. Dees : [Angleterre] ; compo. traits occ. et a.n., ca 1175 ; ms. tour., ca 1238 ; MS2 (ANON. 1963) ; DEAF : *AdamA*.

Néanmoins, dans presque tous ces exemples, le syntagme *mal* + *nom* est précédé d'une préposition : *a mal ure, de mal eure, a mal estrine, de mal aventure, en mal entente, en mal envie, de mal uxor*. Ce n'est pas le cas dans *mal estrine* en (5.52), dans le contexte spécifique d'une malédiction (*donner* + *mal estrenne*), ni de *mal entente* en (5.54). Par ailleurs, en (5.55), la préposition en dans *en mal envie* n'a peut-être pas le même statut que les autres prépositions citées ; si la syntaxe implique que c'est la bonté qui est transformée en mauvais désir, l'usage de la préposition est différent de l'usage circonstanciel des autres occurrences prépositionnelles. Cette construction prépositionnelle est très courante : sur les 1212 attestations de l'adjectif *mal* dans le NCA (lemme *mal* mal), 356 s'insèrent dans une structure *préposition* + *mal* + *nom*, soit 30% des occurrences, alors que cette structure ne concerne que 9% de l'ensemble des adjectifs (14 448 occurrences sur 158 813). Ce n'est pas le cas des deux autres lemmes cités *supra*, « saint » n'apparaissant dans cette structure que dans 9% de ses occurrences, et « lonc » dans 10% de ses occurrences. Aussi, *mal* semble disposer d'un statut spécifique, probablement lié à son sens abstrait, qui génère des possibilités combinatoires étendues, lesquelles se manifestent notamment par de très nombreuses structures prépositionnelles. Au demeurant, au féminin singulier, cette structure *préposition* + *mal* + *nom* est attestée 182 fois, mais ne génère que 17 de nos 31 occurrences de *mal* en défaut de <e>. Elle ne saurait donc être responsable de tous les défauts de <e> de ce lemme ; mais elle est associée à un taux de défaut de <e> légèrement plus élevé (9%) que celui du lemme dans tous les contextes féminins singuliers (6%).

Autres lemmes et défaut de méthode. Nous n'avons pas identifié de collocations récurrentes dans les autres lemmes à défaut de <e>. Une étude exhaustive des collocations nécessiterait la mise en place d'une méthodologie idoine, et l'utilisation d'outils informatiques appropriés. Dans la mesure où ce n'est pas l'objet premier de cette thèse, nous ne proposons pas l'utilisation ou le développement de tels outils, et nous nous en tenons à des remarques ponctuelles comme celles figurant *supra*. Sans de tels outils, il est impossible de trancher dans le continuum sur lequel se placent les collocations, les locutions et les locutions lexicalisées, voire même de dissocier ce qui tient de la simple fréquence relativement élevée et de la collocation. Au demeurant, ces méthodes impliquent des modèles mathématiques complexes qui dépassent la technique de cette thèse, et la question des mesures à utiliser est loin d'être réglée même au sein des études spécialisées (DIVJAK 2019, p. 56-70).

5.2.4 Distribution contextuelle

Les graphiques de séries temporelles en fig. 5.10.a-d, page suivante, représentent le taux de défaut de <e> des 45 lemmes retenus pour chaque texte, devant initiale consonantique (a,c) ou devant initiale vocalique (b,d). Les boîtes à moustaches et les valeurs du test de Student (f,g) permettent la comparaison statistique des deux contextes. Ces figures montrent qu'au-delà du caractère plus ou moins marqué du pôle 1150-1200 évoqué en sec. 5.2.2, et qui coïncide avec l'explosion du nombre de textes telle que représentée en fig. 5.6.b, la principale différence entre les contextes antéconsonantiques et antévocaliques tient au taux global de défaut de <e>. Le taux moyen par texte est de 0,2% ou 0,1% devant initiale consonantique, contre 0,6 ou 0,5% devant initiale vocalique, selon que l'on pondère par le nombre d'occurrences des textes. Si, en chiffres absolus, cette différence semble mineure, elle témoigne d'un effet robuste : les taux de défaut de <e> sont entre 3 et 5 fois plus élevés devant initiale vocalique que devant initiale consonantique. Cet effet est confirmé par le test de Student ($t = 4,84$; $p < 0,005$).³³ Comme l'illustrent les boîtes à moustache en fig. 5.10.e, puisque la moyenne est très faible et que la médiane est à zéro pour les deux contextes, ce sont les troisièmes et quatrièmes quartiles qui font la différence : le troisième quartile dispose encore d'un taux de défaut de <e> nul devant initiale consonantique, alors qu'il englobe des taux positifs devant initiale vocalique, et que le dernier quartile s'étire au-dessus d'approximativement 10% de défaut de <e> devant initiale vocalique.

Les graphiques et mesures en fig. 5.10 témoignent donc de deux effets : d'une part, un effet positif du contexte antévocalique sur le défaut de <e>, et d'autre part un effet plutôt négatif du contexte antéconsonantique sur le pôle 1150-1200, opposé à un effet plutôt positif du contexte antévocalique sur ce même pôle. En revanche, la remontée des régressions locales vers la fin de la période n'est pas consolidée, dans la mesure où elle est principalement due au fait qu'on trouve moins de textes n'ayant aucun défaut de <e> après 1300, mais aussi moins de textes tout simplement.

33: Sauf mention du contraire, nous faisons toujours référence au test à variables indépendantes.

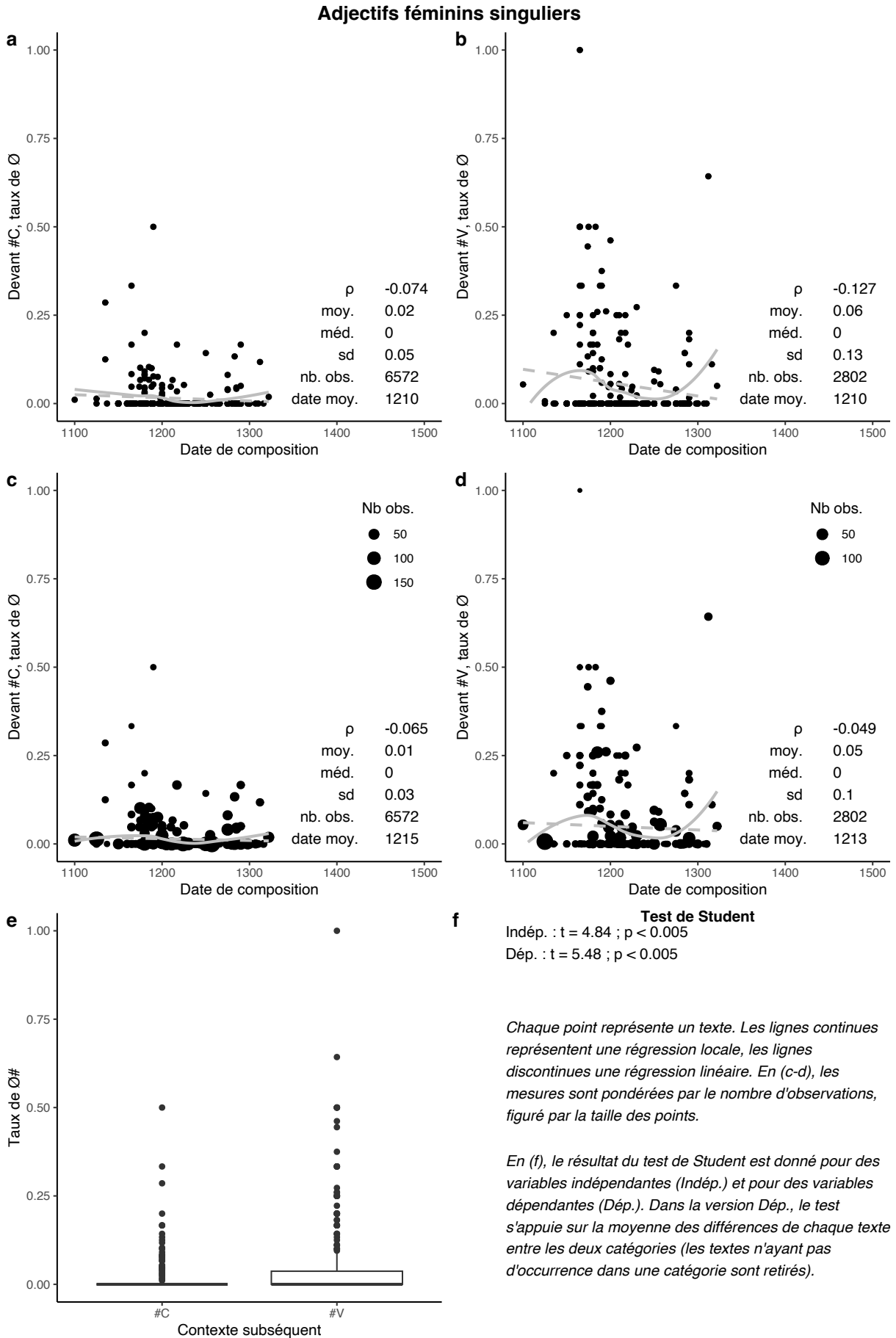


FIGURE 5.10 – Distribution contextuelle des variantes d'adjectifs féminins singuliers en défaut de $\langle e \rangle$.

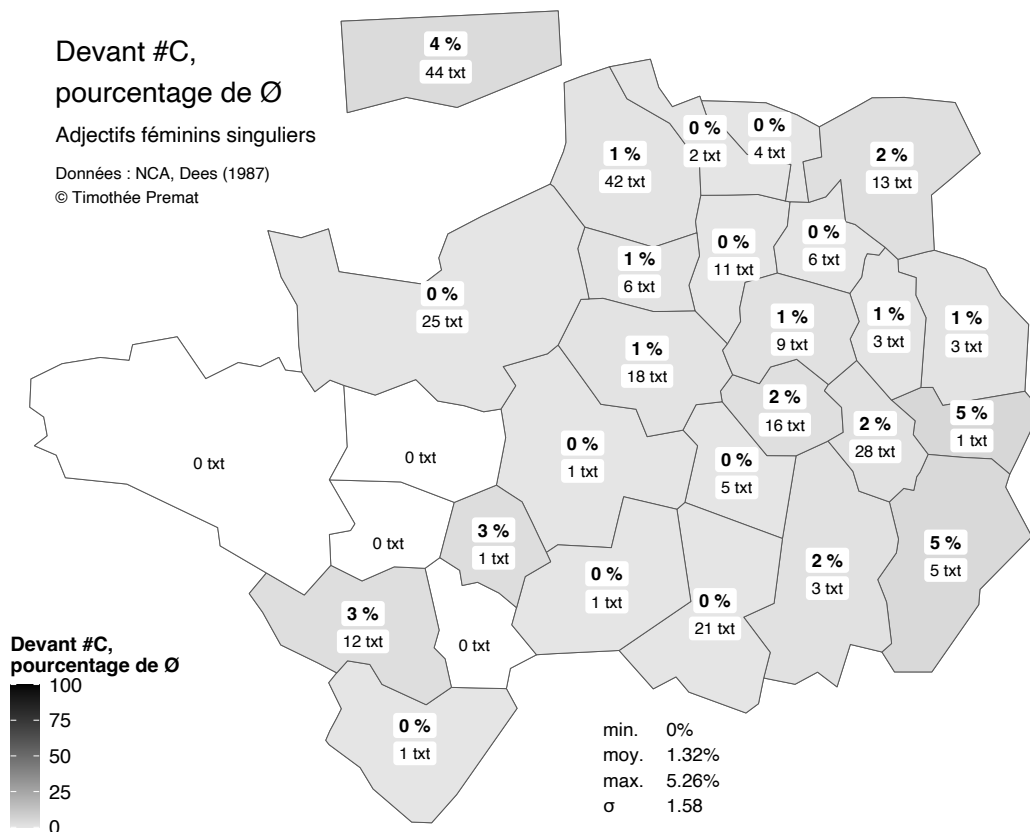
5.2.5 Distribution diatopique

La distribution diatopique de ces défauts de <e> pour les adjectifs féminins singuliers, telle que représentée en fig. 5.11, semble se caractériser comme un bruit de fond constant : si certaines régions peuvent ne montrer de défaut de <e> que devant initiale vocalique (Normandie, Nord, Aisne, Ardennes, Orléanais, région Nièvre, Allier) ou devant initiale consonantique (Indre-et-Loire, Vosges, Oise, Marne, Bourgogne), ce sont, à l'exception de la Normandie, de l'Aisne et de la région Nièvre, Allier, des régions où six textes ou moins répondent à nos requêtes (par contexte). Seules quatre régions répondant à nos requêtes de lemmes ne disposent d'aucune occurrence de défaut de <e> : la Charente-Maritime, la région Indre, Cher, l'Yonne et le Hainaut. Si au moins un de nos lemmes apparaît dans quatre textes (sur 6) dans le Hainaut et dans 5 et 3 textes dans l'Yonne (sur 5), la Charente-Maritime et la région Indre, Cher ne disposent que d'un seul texte (où, par ailleurs, au moins un de nos lemmes apparaît). On peut donc établir que rares sont les régions à échapper à ces défauts de <e>.

Cette composante de *bruit de fond* assez faible (et d'autant plus faible qu'elle ne concerne que 5% des lemmes) semble relativement homogène devant initiale consonantique (écart-type des moyennes des régions : 1,58), avec un maximum à 5,26% de formes en <e> pour nos 45 lemmes atteint en Bourgogne. La situation est moins homogène devant initiale vocalique ($\sigma = 4,56$), avec un maximum à 18,3% pour les textes anglo-normands, et des taux supérieurs au maximum du contexte pré-consonantique (5,26%) en région Vendée, Deux-Sèvres (11%), dans le Nord (10%), en région Moselle, Meurthe-et-Moselle (8%), en Wallonie (7%) et en Normandie (6%). D'autres régions disposent de taux supérieurs au taux régional moyen de défaut de <e> devant initiale vocalique (3,44%), sans qu'ils soient supérieurs au taux maximal devant initiale consonantique : l'Orléanais et l'Aube (5%), ainsi que l'Aisne (4%). Ces taux supérieurs à la moyenne sont consolidés, pour l'anglo-normand et le normand, par un nombre de textes répondant aux requêtes élevé (40 textes en Angleterre, 23 en Normandie) et moyennement élevé dans l'Aube (16 textes), en Wallonie (13 textes), dans l'Aisne (11 textes) et en région Vendée, Deux-Sèvres (10 textes).

Ainsi, des régions se détachent assez clairement du bruit de fond devant initiale vocalique, ce qui n'est pas le cas devant initiale consonantique, et ce sont toujours des régions périphériques. Ce n'est néanmoins pas le cas de toutes les régions périphériques, que ce soit par manque de textes (dans l'absolu ou par manque de textes qui répondent à nos requêtes lemmatiques contextuelles) ou par une présence véritablement marginale des défauts de <e> — ainsi la région Somme, Pas-de-Calais répond bien à nos requêtes (42 textes devant #C, 41 devant #V), mais ne dispose que de 1% et 2% de défaut moyens de <e>. Parmi ces régions marginales qui comportent plus de défauts de <e> que les régions centrales, devant initiale vocalique, l'Angleterre se singularise par un taux de 18%, tandis que la région Vendée, Deux-Sèvres atteint 11% et que tous les autres taux sont inférieurs ou égaux à 10%. On n'observe pas d'autre variable géographique pertinente que cette marginalité des régions; les taux élevés de défaut de <e> devant initiale vocalique étant situés aussi bien au Nord-Ouest (Angleterre, Normandie), au Nord (Nord), au Nord-Est

a



b

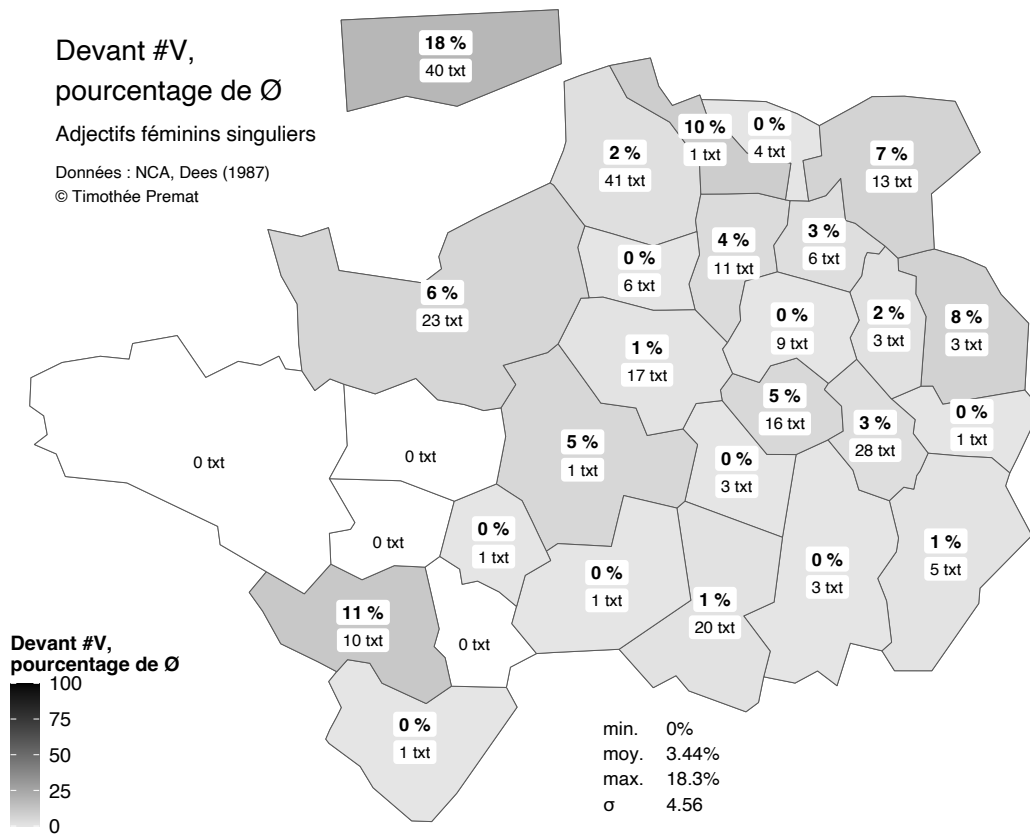


FIGURE 5.11 – Distribution diatopique des alternances <e>/Ø, sur l’ensemble de la période considérée, adjectifs féminins singuliers, suivis d’une initiale consonnante (a) et suivis d’une initiale vocalique (b).

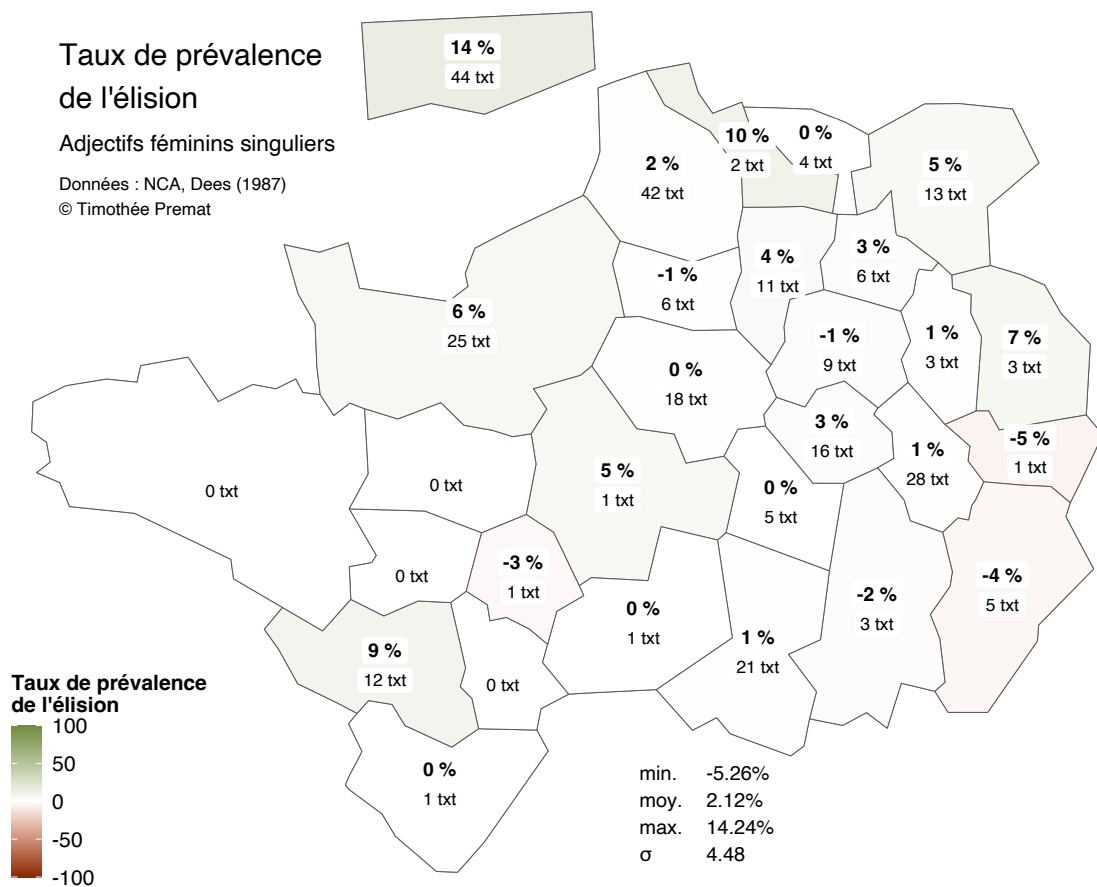


FIGURE 5.12 – Distribution diatopique de la prévalence de l'élision, sur l'ensemble de la période considérée, adjectifs féminins singuliers

(Wallonie, région Moselle, Meurthe-et-Moselle) qu'au Sud-Ouest (région Vendée, Deux-Sèvres); seul le quart Sud-Est semble échapper à cette tendance.

Enfin, la carte de prévalence de l'élision en fig. 5.12 confirme que ces régions marginales sont atteintes par un déséquilibre entre les taux de <e> devant initiale consonantique et ceux devant initiale vocalique : l'indice de prévalence de l'élision (cf. 3.14, p. 81) est supérieur à 4% pour ces régions, avec jusqu'à 14% en Angleterre (4% $_#C$ vs 18% $_#V$), 10% dans le Nord (0% $_#C$ vs 10% $_#V$), 9% en région Vendée, Deux-Sèvres (3% $_#C$ vs 11% $_#V$), 7% en région Moselle, Meurthe-et-Moselle (1% $_#C$ vs 8% $_#V$), 6% en Normandie (0% $_#C$ vs 6% $_#V$) et 5% en Wallonie (2% $_#C$ vs 7% $_#V$) et dans l'Orléanais (0% $_#C$; 5% $_#V$). Ce ne sont pas les seules régions à avoir un indice de prévalence de l'élision supérieur à la moyenne (Aisne : 4%; Aube et Ardennes : 3%; moyenne : 2,12%); cette corrélation entre indice de prévalence de l'élision et taux élevés de formes en défaut de <e> devant initiale vocalique semble donc évoquer une condition nécessaire plus qu'une condition suffisante. Enfin, la seule marge géographique qui ne répondait pas à la tendance repérée *supra*, le quart Sud-Est, dispose d'indices de prévalence de l'élision négatifs :

-5% dans les Vosges (5% __#C vs 0% __#V), -4% en Franche-Comté (5% __#C vs 1% __#V), et -2% en Bourgogne (2% __#C vs 0% __#V). Ces indices négatifs sont dus à un taux de défaut de <e> devant initiale consonantique élevé (le taux de 5% des Vosges et de la Franche-Comté sont le maximum géographique) tandis que les taux de défaut de <e> devant initiale vocalique sont faibles (1% en Franche-Comté) ou nuls (0% en Bourgogne et dans les Vosges).

Il faut noter que le taux de défaut de <e> est toujours un peu moins fiable devant initiale vocalique que devant initiale consonantique, dans la mesure où les initiales consonantiques sont toujours plus nombreuses en a.fr. que les initiales vocaliques; cela se traduit, ici, par 2,5 fois plus d'occurrences (avec ou sans défaut de <e>) devant initiale consonantique que devant initiale vocalique. Néanmoins, cette différence ne saurait, seule, expliquer la disparité entre les taux devant initiales consonantiques et vocaliques, dans la mesure où le nombre d'occurrences devant initiale vocalique est satisfaisant, ce dont témoigne le caractère hautement significatif du test de Student ($p < 0,005$). Ainsi, en écrasant l'ensemble de la période considérée en une synchronie, il faut constater que seules les régions en marge du domaine (sauf au Sud-Est) sont susceptibles de témoigner de forts taux de défaut de <e> devant initiale vocalique, cet effet étant beaucoup moins fort devant initiale consonantique. Ce témoignage est corrélé au taux de prévalence de l'élision, mais cette corrélation ne semble découler que du fait que le contexte préconsonantique empêche les taux de défaut de <e> de s'élever au-dessus de 5,51%.

5.2.6 Interface des distributions diatopiques et diachroniques

Dans cette sous-section, nous interrogeons le recoupement entre diachronie et diatopie *via* des cartes représentant l'une des quatre périodes de notre corpus; ces périodes correspondent à des demi-siècles, sauf la dernière, qui va de 1250 à 1322. Ces cartes prennent d'emblée en compte le contexte mais, pour ne pas multiplier inutilement le nombre de cartes, nous ne reproduisons que les cartes concernant le contexte prévoicalique et celles représentant la prévalence du défaut de <e> devant initiale vocalique sur le défaut de <e> devant initiale vocalique (*prévalence de l'élision*); les résultats pour le contexte préconsonantique sont directement déductibles, pour chaque région, par la soustraction de l'indice de prévalence de l'élision au taux de défaut de <e> devant initiale vocalique.

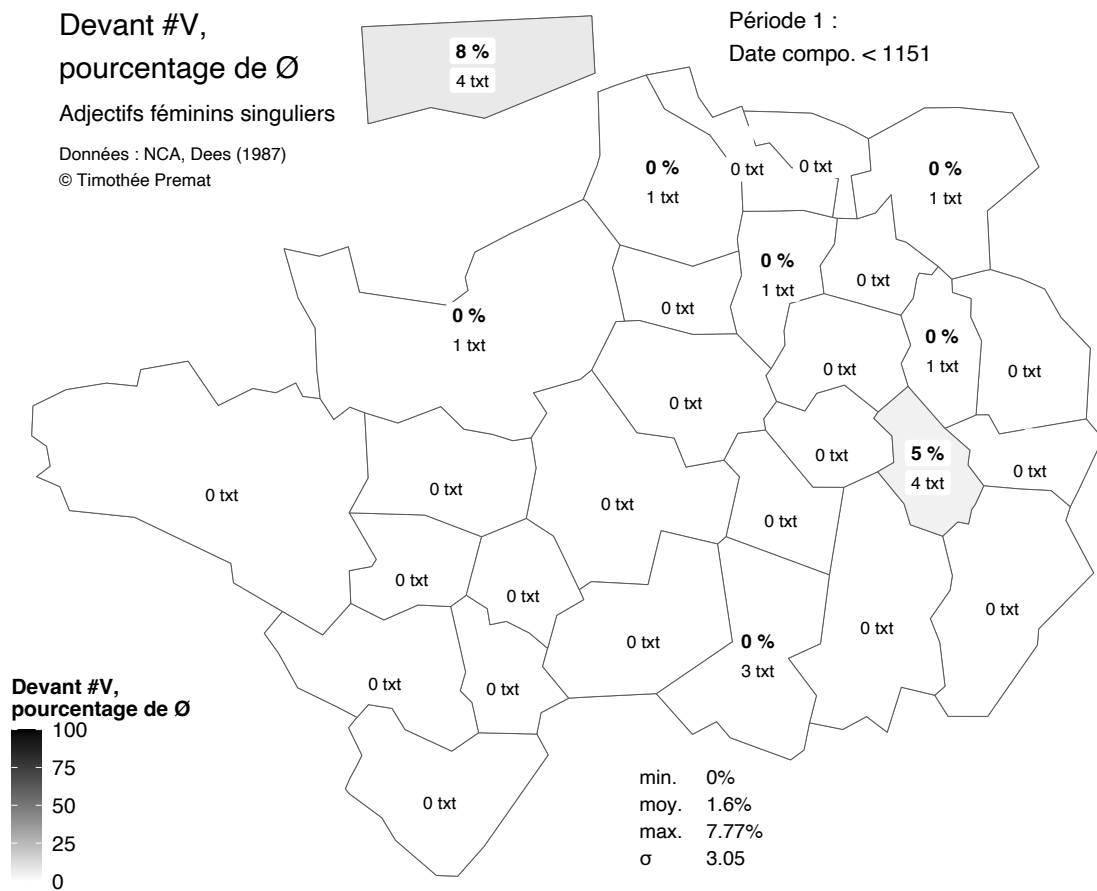


FIGURE 5.13 – Distribution diatopique des alternances <e>/Ø, période 1 (1100-1150), adjectifs féminins singuliers suivis d’une initiale vocalique.

1100 - 1150. Comme indiqué en fig. 5.13 et 5.14, seules deux régions disposent de formes à défaut de <e> antérieures à 1151 : l’Angleterre et la Haute-Marne. Cela n’est pas simplement dû à une absence de formes en défaut de <e> dans les textes ; la plupart des régions n’ont, à cette époque, aucun texte : cinq régions disposent d’un texte (Normandie, région Somme, Pas-de-Calais, Wallonie, Aisne et Meuse), deux disposent de trois textes (Bourgogne et région Nièvre, Allier) et la Haute-Marne et l’Angleterre disposent de quatre textes (fig. 2.9, p. 36. Parmi ces régions, la Normandie, la région Somme, Pas-de-Calais, l’Aisne, la Wallonie, la Meuse et la région Nièvre, Allier n’ont pas de formes en défaut de <e> pour nos 45 lemmes. En revanche, l’Angleterre, et la Haute-Marne ont des occurrences en défaut de <e>. Comme l’indique la carte de prévalence de l’élision en fig. 5.14, ces défauts de <e> se situent principalement devant initiale vocalique en Angleterre (8% __#V, 1% __#C),³⁴ et au contraire plutôt devant initiale consonantique en Haute-Marne (5% __#V, 10% __#C).³⁵

34: Cf. (5.18, 5.19, 5.22) pour roland et (5.23 - 5.25) pour psautier.

35: Cf. (5.29) pour marga et (5.30, 5.31) pour mgrri.

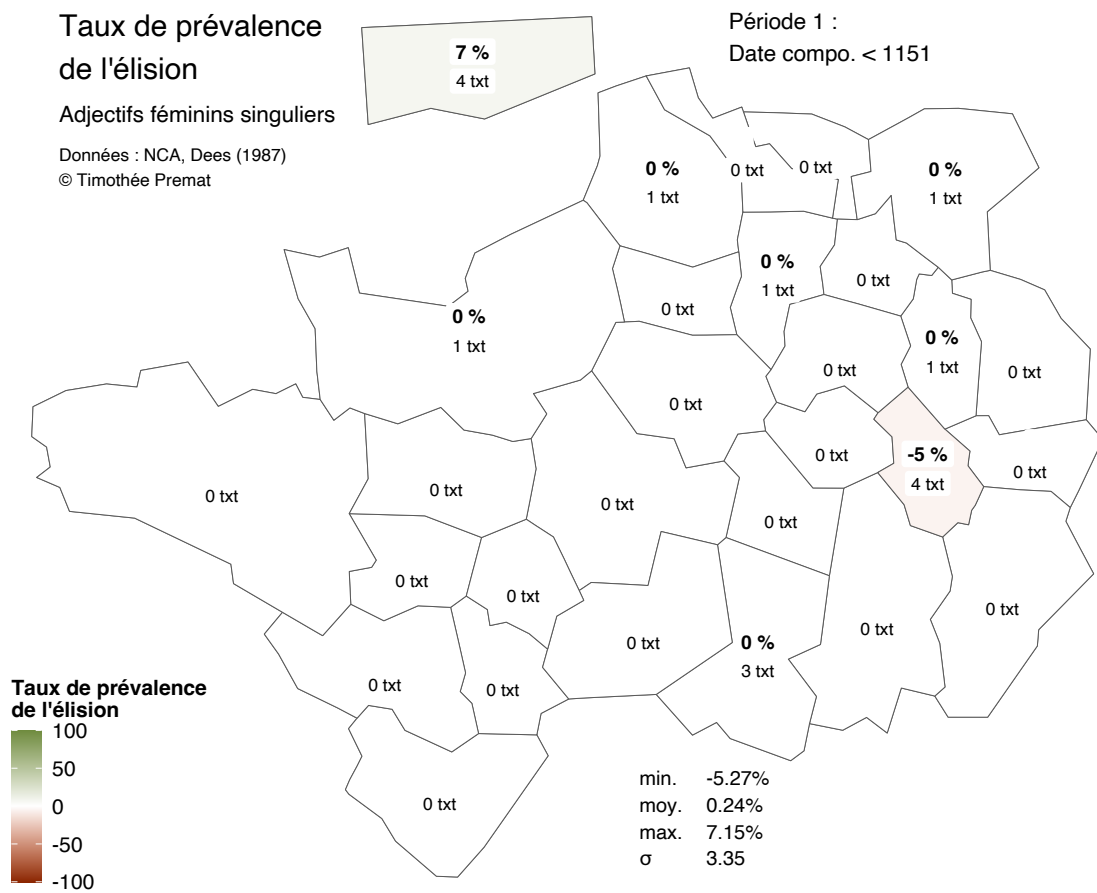


FIGURE 5.14 – Distribution diatopique de la prévalence de l'élision, période 1 (1100-1150), adjectifs féminins singuliers.

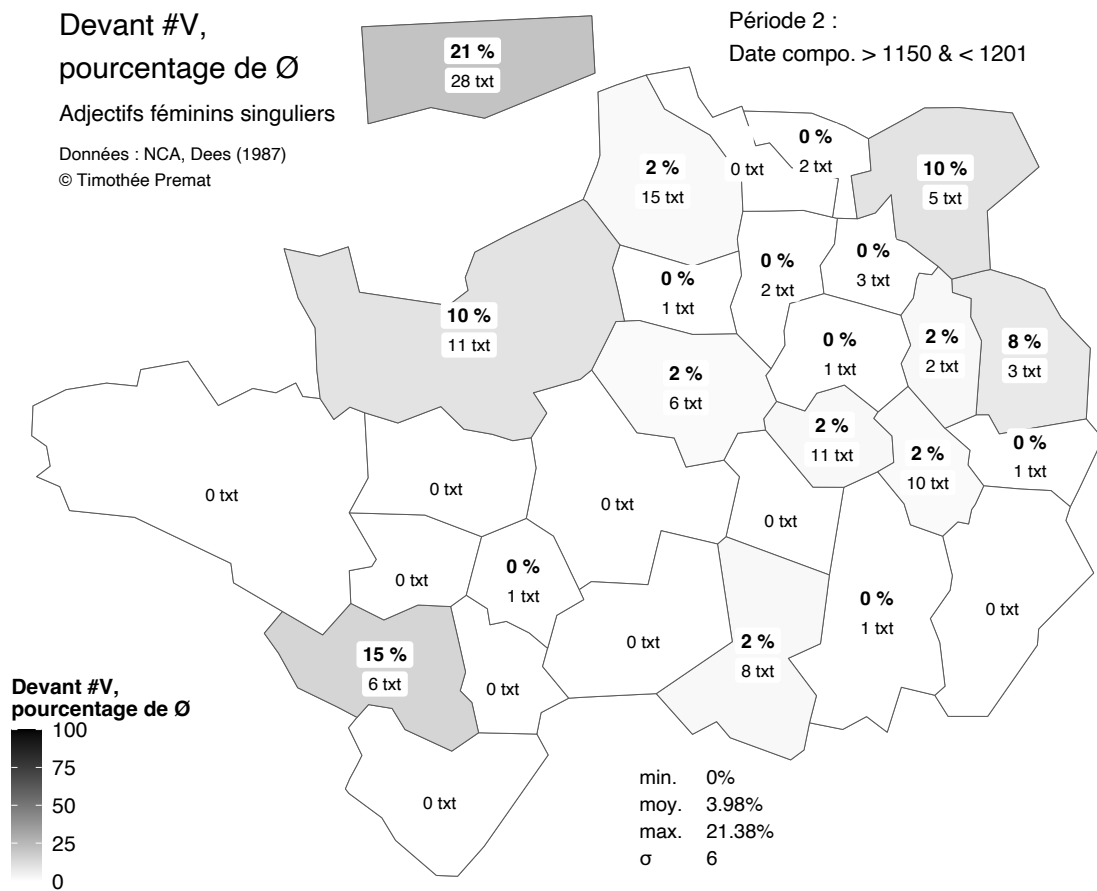


FIGURE 5.15 – Distribution diatopique des alternances <e>/Ø, période 2 (1151-1200), adjectifs féminins singuliers suivis d’une initiale vocalique.

1151-1200. Dans la seconde moitié du XII^e s., les textes se font plus abondants. Ils permettent l’observation de la disposition marginale (sauf au Sud-Est) des taux de défaut de <e> devant initiale vocalique supérieurs à la moyenne, avec 21% en Angleterre, 15% dans la région Vendée, Deux-Sèvres, 10% en Normandie et en Wallonie et 8% en région Moselle, Meurthe-et-Moselle (fig. 5.15). Devant initiale consonantique, on observe au contraire un taux relativement élevé dans les Vosges (5% __#C vs 0% __#V) et, en position géographique non marginale, dans quelques régions plus ou moins au centre du domaine, des taux positifs mais inférieurs à la moyenne (moy. = 4%), avec 3% en Indre-et-Loire (vs 0% __#V) et en région parisienne (vs 2% __#V).

L’indice de prévalence de l’élision, tel que représenté en fig. 5.16, semble déjà préfigurer la situation observée de manière synchronique : dans les marges du domaine, à l’exception du Sud-Est, le défaut de <e> prédomine devant #V, alors qu’au Sud-Est il prédomine devant #C. Par ailleurs, l’Angleterre se signale déjà comme le lieu où le défaut de <e> devant initiale vocalique est le plus marqué (21%); la majorité des textes anglo-normands provient de cette période et, sur les 35 textes anglo-

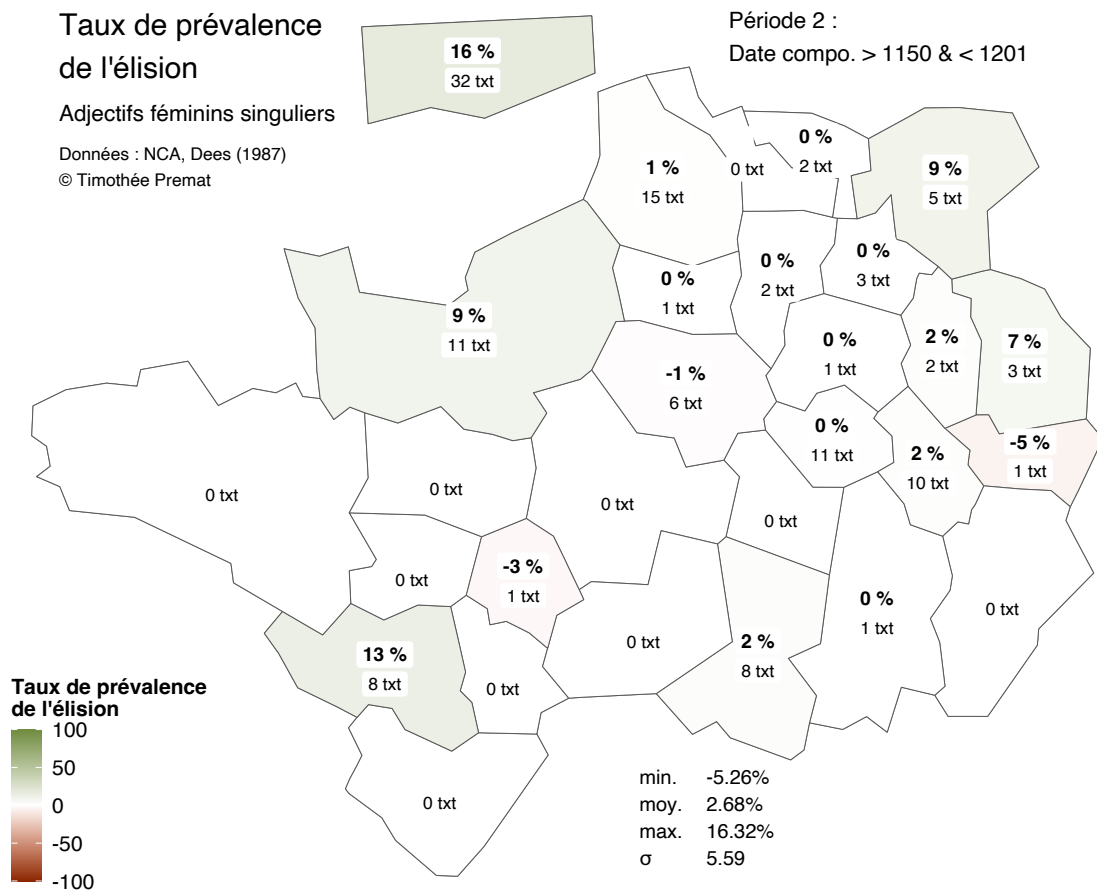


FIGURE 5.16 – Distribution diatopique de la prévalence de l'élision, période 2 (1151-1200), adjectifs féminins singuliers.

normands de cette période, 28 répondent à nos requêtes lemmatiques devant initiale vocalique. Pour les autres marges du domaine indiquant un effet positif du contexte prévocalique sur le défaut de <e>, les taux sont relativement consolidés par les nombres de textes répondant à nos requêtes (11 textes en Normandie, 6 en région Vendée, Deux-Sèvres, 5 en Wallonie); seule la région Moselle, Meurthe-et-Moselle dispose de moins de 5 textes (3 textes). Au contraire, les régions où c'est le contexte préconsonnantique qui semble favoriser le plus les défauts de <e> ne comportent qu'un seul texte répondant à nos requêtes (Vosges et Indre-et-Loire); leurs résultats ne sont donc représentatifs que d'un seul texte et non d'une région en soi.

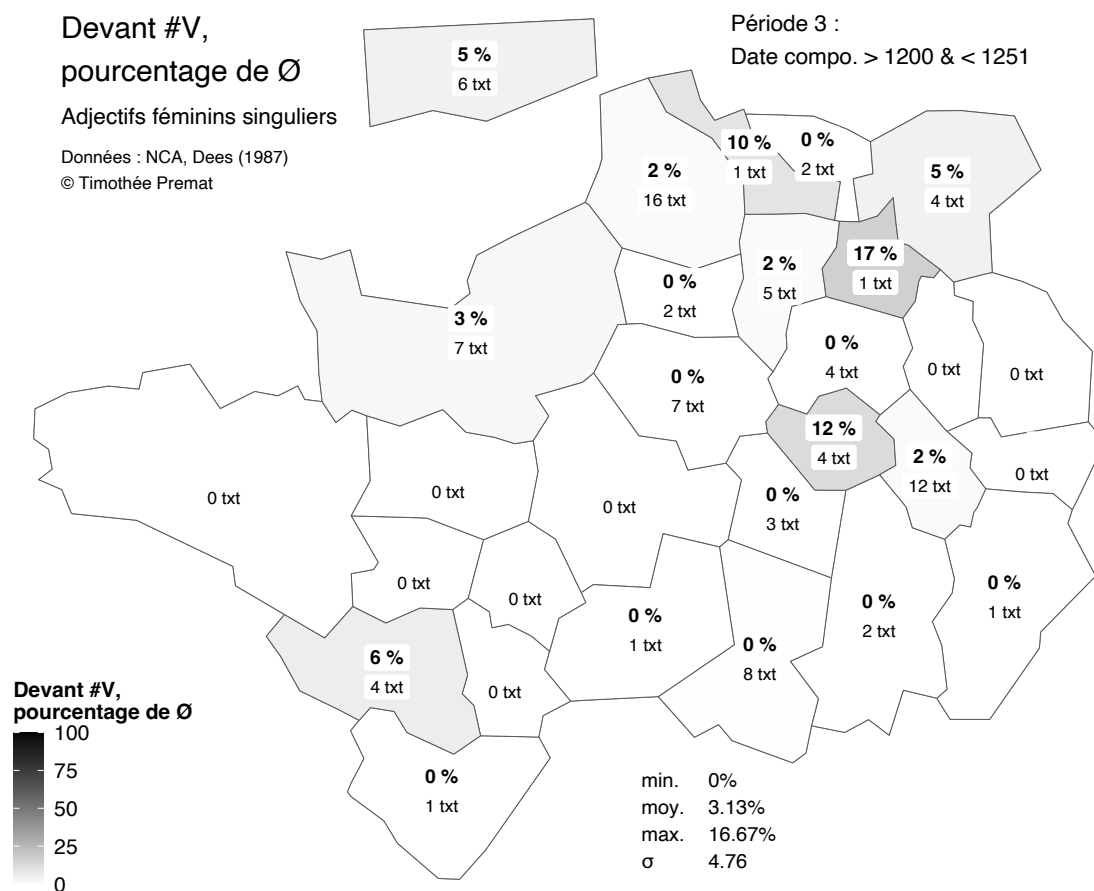


FIGURE 5.17 – Distribution diatopique des alternances <e>/Ø, période 3 (1201-1250), adjectifs féminins singuliers suivis d’une initiale vocalique.

1201-1250. La première moitié du ^{xiii}e s., représentée en fig. 5.17, voit les occurrences en défaut de <e> pénétrer plus fortement à l’intérieur du territoire, avec, devant initiale vocalique, des taux de 17% dans les Ardennes (1 texte) et 12% dans l’Aube (4 textes). Les marges perdent l’ascendant, avec un maximum de 10% dans le Nord (1 texte), puis 6% en région Vendée, Deux-Sèvres (4 textes), 5% en Angleterre (6 textes) et en Wallonie (4 textes), et seulement 3% en Normandie (7 textes), alors que la moyenne des régions s’établit à 3,13%.

Comme représenté en fig. 5.18, la prévalence générale de l’effet du contexte prévocalique sur celui du contexte préconsonantique se maintient (moy. = 2%), tandis qu’au Sud-Est le rapport demeure inversé : c’est le contexte préconsonantique qui semble favoriser le plus l’apparition d’occurrences en défaut de <e> (Franche-Comté : 14% __#C, 0% __#V ; Bourgogne : 2% __#C, 0% __#V), mais ces deux régions demeurent faibles en nombres de textes (Franche-Comté : 1 texte, Bourgogne : 2 textes). On observe donc, dans la première moitié du ^{xiii}e s., un affaiblissement de la spécificité des marges, qui tient non seulement à l’extension à d’autres régions des occurrences en défaut de <e>, mais également à l’affaiblissement relatif

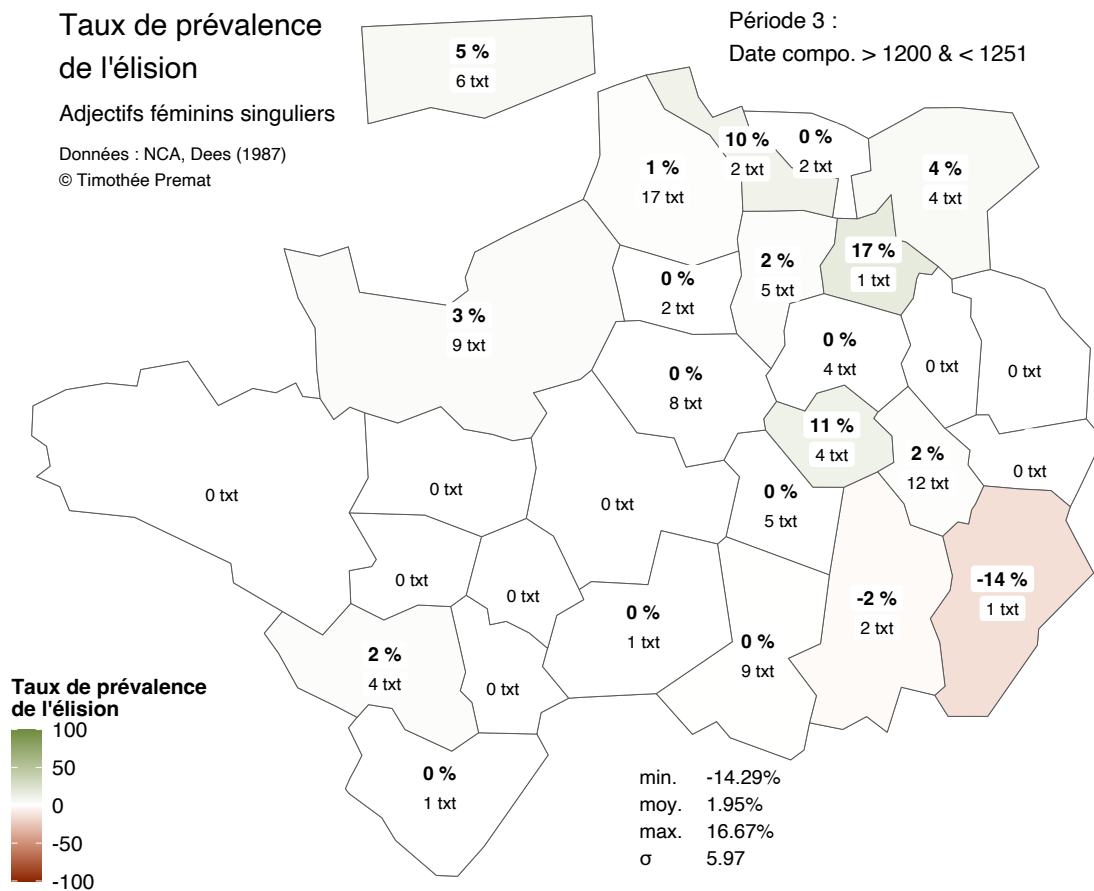


FIGURE 5.18 – Distribution diatopique de la prévalence de l'élision, période 3 (1201-1250), adjectifs féminins singuliers.

des taux de défaut de <e> des régions marginales ; leurs taux demeurent néanmoins supérieurs à la moyenne des taux régionaux (3,13%), sauf en Normandie (3%). Hors du Sud-Est, les taux de défaut de <e> devant initiale consonantique demeurent faibles et rares, avec un maximum de 4% en région Vendée, Deux-Sèvres (*vs* 6% __#V) et 1% dans l'Aube (*vs* 12% __#V), en Wallonie (*vs* 5% __#V) et dans la région Somme, Pas-de-Calais (*vs* 2% __#V). Au sein-même du Sud-Est, seule la Franche-Comté dispose d'un taux considérable de défaut de <e> devant initiale consonantique, avec 14% (*vs* 0% __#V), mais ne contient qu'un seul texte, tandis que le taux de la Bourgogne (2% __#C *vs* 0% __#V) est plus faible et que les Vosges n'ont pas de texte à cette période.

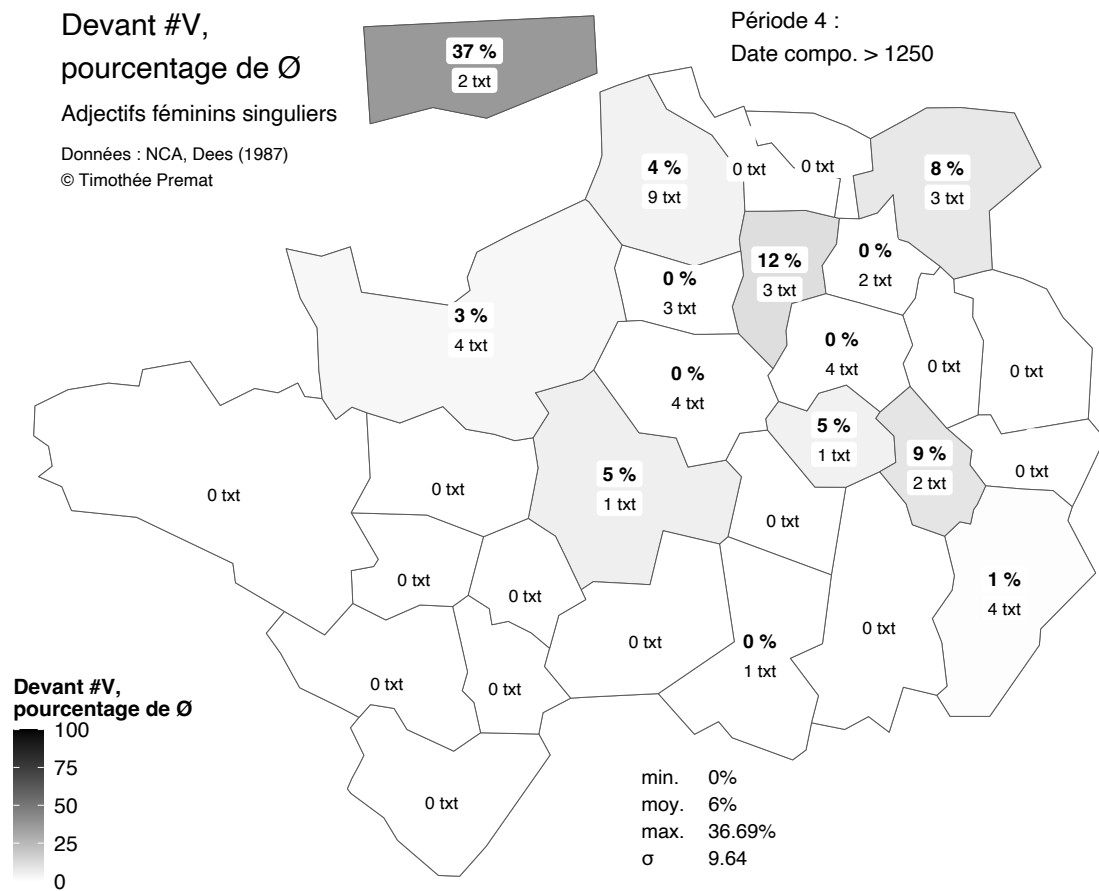


FIGURE 5.19 – Distribution diatopique des alternances <e>/Ø, période 4 (1251-1322), adjectifs féminins singuliers suivis d’une initiale vocalique.

1251-1322. Enfin, dans la dernière partie de notre corpus, qui couvre la seconde moitié du XIII^e s. et le début du XIV^e s., la disparition de la spécificité des marges se confirme. En effet, devant initiale vocalique (fig. 5.19), seules l’Angleterre (37%, 2 textes) et la Wallonie (8%, 3 textes) sont des régions marginales disposant de taux significativement supérieurs à la moyenne (6%), tandis que l’Aisne (12%, 3 textes) et la Haute-Marne (9%, 2 textes) dépassent la Wallonie tout en n’étant pas géographiquement marginales. Aussi, dans cette période, il ne s’agit plus d’une spécificité des marges du domaine, mais d’une spécificité du domaine anglo-normand, tandis que des régions de la moitié septentrionale du domaine continental peuvent disposer de taux élevés de défaut de <e> sans être marginales. Néanmoins, cette disparition du caractère marqué des marges est aussi due à la disparition d’un large quart Sud-Ouest du domaine, qui ne dispose plus d’aucun texte dans le corpus.³⁶ Parallèlement, la Normandie et la région Somme, Pas-de-Calais conservent des attestations de défaut de <e> devant initiale vocalique, mais à des taux inférieurs à la moyenne (respectivement 3% sur 4 textes et 4% sur 9 textes); ces régions n’ont pas d’occurrences de défaut de <e> devant initiale consonantique. Ainsi, le contexte prévocalique continue à l’emporter dans la plupart des régions

36: Cf. fig. 2.10, p. 37.

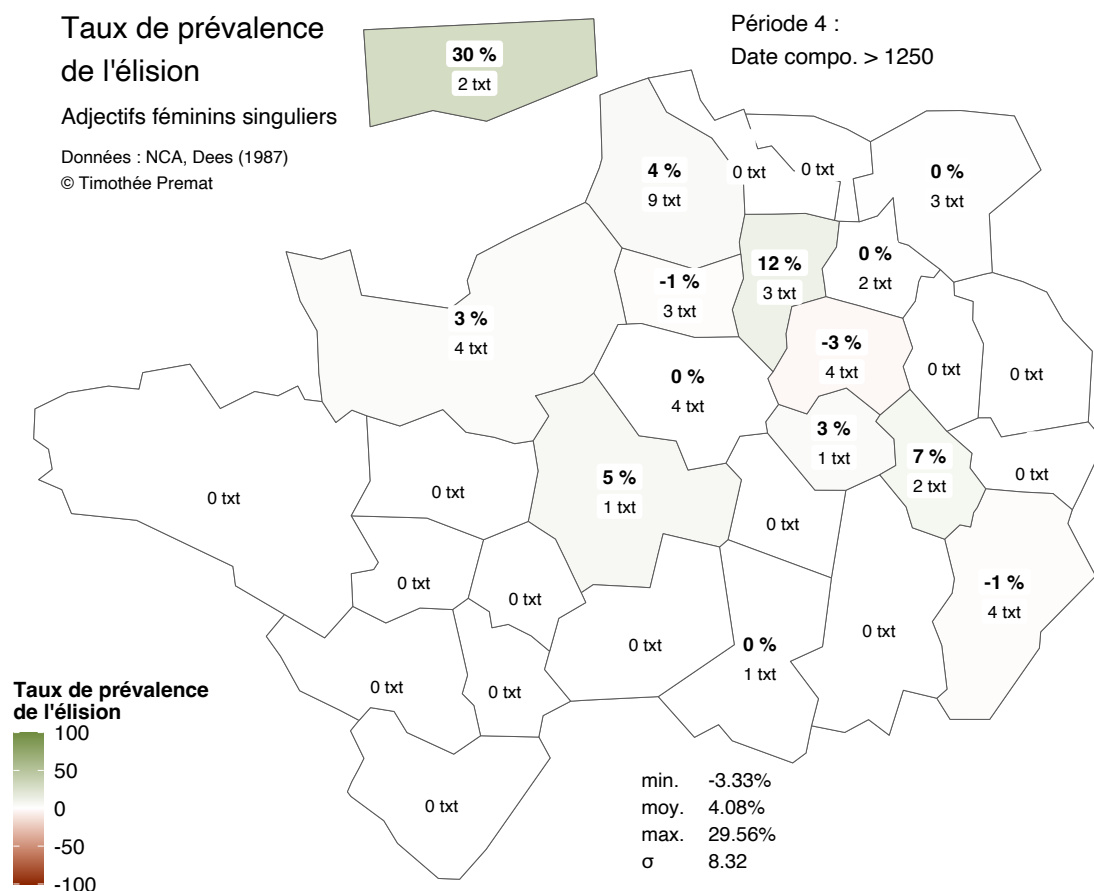


FIGURE 5.20 – Distribution diatopique de la prévalence de l'élision, période 4 (1251-1322), adjectifs féminins singuliers.

disposant d'occurrences de défaut de <e> (fig. 5.20) ; le contexte préconsonantique n'a un effet plus prononcé que le contexte prévocalique que dans la Marne (3% __#C, 0% __#V, 4 textes), l'Oise (1% __#C, 0% __#V, 3 textes) et la Franche-Comté (2% __#C, 1% __#V, 4 textes). Lorsque c'est le contexte préconsonantique qui est le plus souvent associé avec des défauts de <e>, l'effet de ce contexte reste donc faible ; alors que dans les régions où c'est le contexte prévocalique qui est le plus souvent associé avec des défauts de <e>, l'effet de ce contexte peut être fort : l'indice de prévalence de l'élision atteint 30% en Angleterre (37% __#V, 7% __#C, 2 textes), et va de 3 à 12% pour les autres régions. Enfin, il faut constater que si la spécificité des marges s'efface tout à fait à cette période, c'est aussi parce qu'une bonne partie des régions marginales qui étaient spécifiques aux périodes antérieures deviennent muettes : le corpus n'a plus aucun texte dans la région Vendée, Deux-Sèvres, en Meurthe-et-Moselle ou dans le Nord.

5.3 Des observations aux généralisations

Plusieurs éléments ressortent des données discutées ci-dessus. Dans cette section, nous entendons passer de l'observation des données à leur analyse, et proposer ainsi une montée en généralité. Néanmoins, certains effets repérés ci-dessus demandent à être explorés plus en détail, ainsi, nous reprenons l'analyse des effets de fréquence en nous appuyant plus précisément sur la bibliographie, ou traitons plus en profondeur certaines dimensions de la diatopie et, en particulier, la question du statut et du poids de l'anglo-normand dans nos données. Dans ce qui suit, nous synthétisons la distribution des taux de défaut de <e> en fonction des lemmes (sec. 5.3.1), du contexte (sec. 5.3.2), de la variable diatopique qui semble s'exprimer sous la forme d'une opposition partielle entre centre et périphérie (sec. 5.3.4), puis nous analysons plus en détail la situation de l'anglo-normand et son poids dans les données (sec. 5.3.5), avant d'interroger globalement la place des défauts de <e> au sein de l'ensemble des adjectifs féminins singuliers (sec. 5.3.6). À chaque fois, nous proposons de résumer l'analyse sous la forme d'une ou plusieurs généralisations, dont le grain, certes grossier, permet la synthèse de l'effet du ou des facteurs impliqués dans la variation <e>/∅.

5.3.1 Lemmes atteints par le défaut de <e>

Comme exposé en sec. 5.2.3, les premières occurrences en défaut de <e> sont présentes dès le premier texte du corpus, et la quantité de lemmes disposant de formes en défaut de <e> augmente parallèlement à la taille du corpus et donc au nombre d'attestations de ces lemmes. Il ne semble donc pas possible d'analyser un phénomène de diffusion lexicale de l'innovation que représenterait le défaut de <e>, dans la mesure où plus de la moitié de nos lemmes sont attestés en défaut de <e> dans la seconde moitié du *xii^e* s. Par ailleurs, nous ne sommes visiblement pas dans le cas d'une innovation linguistique en cours de généralisation, sa productivité culminant dans la seconde moitié du *xii^e* s. et décroissant ensuite.

(5.57) **ADJFÉMSG : diffusion lexicale.**

Il n'y a pas de phénomène de diffusion lexicale du défaut de <e> observable.

Par ailleurs, la corrélation entre fréquence des lemmes et taux de défaut de <e> de ces lemmes est faible ($\rho = -0,29$) et non significative ($p = 0,44$). La constatation du profil décroissant de la régression linéaire en fig. 5.7 va à l'encontre des prédictions que l'on pourrait extrapoler des études sur les effets de la fréquence. Typiquement, plus un item est fréquent, plus il est sujet au REDUCING EFFECT, qui consiste en une suppression accentuée de certains segments (BYBEE 2006, p. 714-715, 724-725, 2000, 2002; BYBEE et SCHEIBMAN 1999). Loin de réfuter ce constat bien acquis dans les études fréquentielles, nos données indiquent plus probablement deux éléments. Le premier élément tient au fait que la fréquence qui conditionne le REDUCING EFFECT est la fréquence d'exposition et de production réelle, alors que nous traitons ici d'un corpus écrit, fragmentaire et concernant une société où l'oralité dominait bien plus qu'elle ne domine les sociétés contemporaines.

Le second tient probablement à l'activation d'un biais du corpus : ses dimensions, conséquentes mais par définition limitées, peuvent filtrer un certain nombre de résultats, avec pour ce filtre une dimension stochastique : les lemmes peu attestés peuvent, à l'issue de ce filtre, être dotés de taux aberrants. Nous avons suggéré que c'était de ce filtre dont témoignait le profil en hyperbole inversée de la régression locale en fig. 5.7. Ce filtre agirait de la manière suivante : si l'on considère qu'un lemme peut disposer de *quelques* occurrences en défaut de <e> (p.ex. 2%, comme « bon » : 34 défauts sur 1405 occ.), la réduction, par échantillonnage, de sa quantité absolue d'occurrences entraîne l'une des deux solutions suivantes : (a) le lemme perd toute attestation de défaut de <e> — c'est la solution la plus probable, ou (b) le lemme conserve quelques attestations en défaut de <e>. Dans la seconde solution, certes moins probable, son taux de défaut de <e> a d'autant plus de probabilité d'être élevé que sa fréquence est basse, selon un profil exponentiel. Si la première solution est plus probable que la première, on peut imaginer qu'elle concerne de nombreux lemmes à faible fréquence, trop faible pour qu'ils exhibent des formes à défaut de <e> (p.ex. $F < 41$ occ., si l'on prend comme repère le taux de 2% de « bon » sur 1405 occ.)³⁷ ; de fait, ces lemmes, sans défaut de <e>, n'apparaissent pas dans nos données. À l'inverse, les quelques lemmes à faible fréquence mais à taux élevés de défaut de <e> témoigneraient de la dimension stochastique d'un tel échantillonnage, qui aurait retenu des formes à défaut de <e>. Si l'on admet cette hypothèse, cet effet n'est pas un effet linguistique, mais est dû à un biais dans le corpus, qui ne comporte (i) pas assez d'occurrences des lemmes concernés par (a) pour attester de leurs défauts de <e>, et qui ne comporte (ii) pas assez d'occurrences des lemmes à fort taux de défaut de <e> pour relativiser leur taux. Ainsi, les six lemmes dont les taux de défaut de <e> sont supérieurs à 10% disposent, au maximum, de 16 occurrences pour l'un d'entre eux, et de moins de 10 occurrences pour les cinq autres. La fréquence de ces six lemmes est donc bien en-deçà de la fréquence qui permettrait, sous l'hypothèse la plus probable (a), au lemme « bon » d'attester d'occurrences en défaut de <e>. Ce point est résumé sous la forme de la généralisation en 5.58.

(5.58) **ADJFÉM_{sg} : fréquence.**

La corrélation entre taux de défaut de <e> et fréquence des lemmes est faible, non significative et inverse. Elle représente un effet de filtre quantitatif, et non un effet linguistique.

($\rho = -0,29$; $p = 0,44$)

Enfin, nous avons observé un troisième phénomène fréquentiel dans l'étude de nos lemmes à défaut de <e> : trois d'entre eux, « saint », « lonc » et « mal » voient leurs défauts de <e> associés à l'inclusion de l'adjectif dans une collocation avec le nom suivant. Ces associations sont plus ou moins restrictives sur le plan de la liberté paradigmatique du second élément,³⁸ mais tous les degrés de grammaticalisation indiqués en (5.59) doivent être considérés comme des *constructions*. La liste en (5.59) est une reprise par Bybee de GOLDBERG (1995, 2003), à laquelle elle ajoute les *prefabs*, c'est-à-dire des associations syntaxiques de classes grammaticales et non d'items lexicaux, que nous numérotions [v] ici.

37: Si « bon » a 34 occurrences en défaut de <e> sur 1405 occurrences, un échantillonnage fidèle comporterait moins d'une occurrence en défaut de <e> en-deçà de 41 occurrences :

$$\frac{34}{1405} \times 41 = 0,99$$

38: Ici, nous ne parlons de la liberté que pour le second élément, puisque notre étude a pour entrée un premier élément donné : le lemme adjectival.

GOLDBERG (1995). *Constructions: A construction grammar approach to argument structure*. — (2003). « Constructions: A new theoretical approach to language ».

(5.59) Degrés de grammaticalisation

- i idioms with fixed lexical content : *go great guns* ;
- ii idioms that are partially filled : *jog <someone's> memory* ;
- iii constructions with some fixed material : *he made his way through the crowd* ; and
- iv fully abstract constructions : *they gave him an award* [...]
- [v] Prefabs can also be considered to be instances of constructions that are lexically filled.

(BYBEE 2006, p. 716)

Les associations que nous avons relevées correspondent à plusieurs de ces types. Certaines semblent tenir au plus haut degré de grammaticalisation (i), comme *mal(e) eure* et *mal(e) aventure* ; les deux derniers sont par ailleurs attestés comme un seul mot par certains éditeurs et lexicographes (*maleur*, *malaventure*). Pour d'autres collocations, il est difficile de définir leur type précis : si *saint(e) eglise* peut être considéré comme une locution à contenu lexical fixe (i), elle représente alors un sous-type de locutions partiellement remplies, comme *saint(e) + nom de sainte*, qui est lui-même un sous-type de construction préfabriquées (v) (*saint + nom*). De même, si la locution [*de*] *longue haleine* est pleinement lexicalisée en fr.mod., son rapport à la longueur du souffle dans un cor dans les occurrences de *Roland* semble indiquer que ce n'est pas encore le cas en a.fr. En revanche, les occurrences comme *long espee* ou *mal entente* semblent ne correspondre qu'au type le moins grammaticalisé, le préfabriqué (v) étant simplement l'expression d'une configuration syntaxique récurrente dont les items ne sont bornés que par leur classe grammaticale.

Plus une construction est grammaticalisée (polarité (i) du continuum en 5.59), plus elle subit deux effets : l'*autonomie* et la *fusion des constituants* (BYBEE 2006, p. 715). Le résultat le plus fort de la grammaticalisation est la réduction du *type* (i.e. syntagme récurrent) à un simple *token* (fusion des constituants : une seule entrée lexicale). Dans une collocation réduite à un seul *token*, les constituants sont déçus de leur statut de *token*, ce qui les rend donc autonomes, sémantiquement et morphologiquement, vis-à-vis des autres instances de leur lemme (il n'y a plus de relation d'analogie entre les mots du *token* et les instances isolées du lemme). Ainsi, le *mal* de *mal(e) eure* n'est plus équivalent au *mal* de *mal(e) uxor*.

En ce qui nous concerne, et au-delà des effets de réduction, et d'autonomie, la fusion des constituants a une conséquence phonologique directe : abolissant des frontières de mots dans le traitement mental des unités grammaticalisées, elle aboutit à la formation d'un hiatus interne lorsque le mot suivant commence par une voyelle : *male + eure*, *longe + aleine*, *seinte + eglise*. Dans la mesure où le contexte prévocalique semble favoriser les défauts de <e> en position finale (sec. 5.3.2), et où les séquences /əV/ interne sont sujettes à réduction des hiatus (cf. chap. 11), un degré plus ou moins fort de grammaticalisation entre un adjectif paroxyton et un nom à initiale vocalique subséquent a toutes les raisons de déclencher un défaut de <e>. Quant à l'autonomie, elle peut éventuellement protéger l'adjectif de la restauration du <e> défectueux, dans la mesure où l'adjectif pris dans une construction grammaticalisée n'est plus mis en série avec les autres formes de son lemme.³⁹ Il est sur ce point notable que, dans nos données, tous les défauts de <e> que nous avons relevés dans des

39: Ainsi, c'est probablement l'autonomie due à la lexicalisation qui a protégé le caractère épïcène de l'adjectif *grand* dans *grand-mère*, attesté dès 1270 par REY (2012). Sans le pouvoir de l'autonomie, on aurait attendu *grande mère*, l'adjectif devenant progressivement fléchi en genre par analogie au type dominant. Ici, l'aboutissement fr.mod. *malheur* témoigne peut-être d'un effet similaire, si ce n'est que c'est une forme défectueuse (*mal* FÉM.sg) qui a été protégée de la restauration de sa flexion ; mais l'hiatus interne impose une interférence considérable au phénomène, le <h> de *heure* n'étant pas aspiré (*l'heure* vs *la hache*).

collocations plus ou moins grammaticalisées sont toujours placés devant une initiale vocalique, sous l'effet probablement conjoint, d'une part du défaut de <e> devant initiale vocalique et de la réduction des hiatus internes (ce qui explique les occurrences relevées), et d'autre part de l'absence ou de la rareté des syncopes en position préconsonantique interne, sauf dans quelques environnements restreints (cf. chap. 11).

Ainsi, puisque les locutions nécessitent une représentation mentale qui leur soit propre et qui ne soit pas réductible à la somme des représentations des items qui les composent (BYBEE 2006, p. 721), il y a lieu de leur concéder un statut spécifique. Dans nos données, trois lemmes apparaissent en défaut de <e> dans des collocations, avec une certaine régularité, mais leur niveau de grammaticalisation va de la simple structure syntaxique récurrente (*prefab*) à l'expression pleinement grammaticalisée, résumable à un seul constituant indivisible. Dans tous ces cas, le défaut de <e> n'apparaît que devant initiale vocalique. Néanmoins, cet élément d'explication n'est en aucun cas suffisant : nous n'avons pas repéré de collocations spécifiques pour la plupart de nos lemmes, et même pour les trois lemmes qui montrent une certaine affinité pour les collocations, celles-ci montrent un degré de grammaticalisation variable (*mal eure* > *mal uxor*; *long aleine* > *long espee*; *seint eglise* > *seint agnes*) et, dans tous les cas, une concurrence avec des formes sans défaut de <e> (*male eure*, *longe aleine*, etc.). Ces éléments peuvent être résumés à la généralisation en (5.60).

(5.60) **ADJFÉMSG : collocations.**

Un défaut de <e> semble être encouragé par la collocation avec un nom à initiale vocalique.

Cela ne concerne qu'une minorité de lemmes et des instances des collocations en question.

Ce point sera central dans l'analyse que nous proposerons dans la dernière partie de ce travail. Alors que certains mots fonctionnels, atteints par la cliticité, sont sujets à l'élision, les mots lexicaux ne l'admettent que lorsqu'ils entretiennent un rapport syntaxique étroit avec le mot suivant, voire sont considérés comme ne formant qu'un seul *token* avec celui-ci. On voit donc ici se dessiner, *via* les collocations, l'un des domaines de l'élision.

5.3.2 Le rôle du contexte

Comme indiqué en sec. 5.2.4 et en fig. 5.10, le contexte prévocalique est associé à des taux de défaut de <e> plus élevés que le contexte préconsonantique. Cet effet est relativement marqué ($t = 4,84$) et est hautement significatif ($p < 0,005$). L'effet est plus fort si l'on s'en tient aux textes qui comportent des occurrences dans les deux contextes (test de *Student* à variables dépendantes : $t = 5,48$; $p < 0,005$). S'il vaut pour l'ensemble des textes, cet effet du contexte prévocalique s'applique néanmoins différemment selon les régions et les périodes. La comparaison du dessin des courbes de régression locale en fig. 5.10 permet ainsi d'observer que le taux moyen de défaut de <e> est bien plus élevé devant initiale vocalique que devant initiale consonantique dans la

seconde moitié du XII^e s. et à la fin du corpus, mais que les deux taux se rapprochent au XIII^e s.

Par ailleurs, comme l'indiquent les cartes de prévalence de l'éliision (synchronique en fig. 5.12, diachroniques en fig. 5.14, 5.16, 5.18 et 5.20), cet effet connaît aussi une variation diatopique marquée : il est inversé dans le Sud-Est du domaine, se concentre aux marges du domaine dans la seconde moitié du XII^e s., avant que ces marges ne se diluent au XIII^e s. Enfin, l'anglo-normand joue un rôle particulier dans cet effet, contribuant en bonne partie au pic de la régression locale dans la seconde moitié du XII^e s. (cf. sec. 5.3.5 *infra*). Ces remarques peuvent être rassemblées en la généralisation (5.61).

(5.61) **ADJFÉMSG : contexte.**

Le contexte prévocalique est associé à plus de défauts de <e> que le contexte préconsonantique.

($t = 4,99$; $p < 0,005$)

L'effet n'est pas stable dans le temps et l'espace.

5.3.3 Diachronie générale

Nous avons pu observer en sec. 5.2.2 une certaine concentration de textes à degré plus élevé de défaut de <e> dans la seconde moitié du XII^e s. Un examen du comportement de chaque contexte a permis de réduire cette tendance à un effet très modéré ou absent devant initiale consonantique (selon que les taux des textes sont pondérés par leur nombre d'occurrences), mais plus marqué devant initiale vocalique. Nous proposerons *infra* que cet effet est dû à une interférence diatopique, causée par l'expression de la singularité des marges (qui peuvent disposer de taux plus élevés que le centre du domaine; sec. 5.3.4) et plus spécifiquement par la présence marquée de textes anglo-normands à cette période (sec. 5.3.5). Cet effet de pic des taux de défaut de <e> devant initiale vocalique dans la seconde moitié du XII^e s. induit une légère décroissance globale des taux de défaut de <e>. Ainsi, en fig. 5.10, la corrélation entre date de composition et taux de défaut de <e> est négative ($\rho = -0,07$ devant initiale consonantique et $-0,13$ devant initiale vocalique). Ce coefficient indique une corrélation très faible, puisque l'évolution diachronique n'a rien, ici, d'une corrélation linéaire.⁴⁰ Nous proposons de résumer ce constat sous la forme de la généralisation (5.62), qui fait elle-même appel à des généralisations avancées *infra*.

(5.62) **ADJFÉMSG : diachronie.**

On observe une concentration de taux de défaut de <e> plus élevés dans la seconde moitié du XII^es.

Cet effet n'est pas homogène sur l'ensemble du territoire (5.63, 5.67).

40: Étant donnée l'intense variance de nos données, le coefficient de corrélation linéaire de Pearson est toujours faiblement satisfaisant (proche de 0), et sa valeur p souvent bien trop élevée (ici $p = 0,21$ devant initiale consonantique). Dans l'usage peu orthodoxe que nous faisons de ce coefficient ici, celui-ci nous sert simplement à estimer, globalement, si les taux de défaut de <e> vont croissant ($\rho > 0$) ou décroissant ($\rho < 0$).

5.3.4 Centre - périphérie

Sur l'ensemble de la période, écrasée en une synchronie artificielle (fig. 5.11), les marges se signalent par une certaine spécificité devant initiale vocalique : elles regroupent les taux les plus élevés (sec. 5.3.4).

Cet effet ne concerne pas toutes les régions marginales : la région Somme, Pas-de-Calais, le Hainaut, le quart Sud-Est ainsi que la marge Sud disposent de taux de défaut de <e> devant initiale vocalique inférieurs à la moyenne, voire nuls. Cet effet est également moins marqué devant initiale consonantique, avec une plus grande proximité entre les taux élevés de l'intérieur du domaine (jusqu'à 3%) et ceux des marges hors Sud-Est (jusqu'à 4%), tandis que le Sud-Est dispose des taux les plus élevés, avec 6% en Bourgogne et 5% en Franche-Comté et dans les Vosges. Ainsi, l'écart type est beaucoup plus élevé devant initiale vocalique ($\sigma = 4,56$) que devant initiale consonantique ($\sigma = 1,58$), traduisant une plus grande hétérogénéité des régions dans le premier cas.

Diachroniquement (sec. 5.2.6), cet effet de contraste entre le centre et la périphérie ne peut s'observer dans la première moitié du XII^e s., par manque de textes. Il se retrouve fortement dans la seconde moitié du siècle, avant de disparaître au XIII^e s., sous l'effet conjoint de l'apparition éparse de taux élevés au centre du domaine et de l'affaiblissement de certaines valeurs extrêmes des marges. Le caractère réfractaire du Sud-Est au défaut de <e> devant initiale vocalique est confirmé dans toutes les périodes. Hormis dans la première partie du XIII^e s., l'anglo-normand se signale par des taux bien plus élevés que la moyenne des régions, mais avec un fort écart entre les périodes. Étant donné le fort nombre de textes anglo-normands du corpus (48 textes), il paraît nécessaire d'étudier plus en détail le comportement des textes de ce domaine, notamment pour comprendre le poids de l'anglo-normand au sein de la diachronie générale du corpus. Avant d'aborder la situation spécifique de l'anglo-normand, ces observations sur la relation entre centre et périphérie peuvent être synthétisées sous la forme de la généralisation (5.63).

(5.63) **ADJFÉMSG : diatopie.**

Les marges comportent des taux de défaut de <e> devant initiale vocalique plus élevés que le centre du domaine.
L'effet n'est pas stable dans le temps et l'espace.

5.3.5 Poids de l'anglo-normand et de son hétérogénéité

L'influence de la diatopie sur le tracé diachronique général ne se résume pas à cette contamination de certaines parties du centre du domaine par ses marges, suivi de l'affaiblissement du contraste entre les marges et centre du domaine. Le facteur qui semble le plus important, ici, tient plus spécifiquement à l'anglo-normand. Non seulement l'anglo-normand dispose, sur l'ensemble de la période (fig. 5.11) et dans deux périodes sur quatre (fig. 5.15 et 5.19), de taux de défaut de <e> devant initiale vocalique bien plus élevés que les autres régions, mais il dispose aussi d'une variation interne sans commune mesure avec celle des autres régions.

L'hétérogénéité de l'anglo-normand peut être approchée par une simple comparaison entre l'écart-type des taux de défaut de <e> en anglo-normand (5.64 a) et en oïl continental (5.64 b); l'écart-type de l'anglo-normand est trois fois plus élevé que celui des dialectes continentaux. Ce caractère élevé de l'écart-type n'est pas un simple effet du nombre de textes; la deuxième région à contenir le plus de textes, la région Somme,

Pas-de-Calais (45 textes), dispose d'un écart-type inférieur à celui de l'ensemble du domaine continental (5.64 c).

(5.64) Écart-types devant initiale vocalique (toute la période).

- (a) A.n. : moy. = 0,18; σ = 0,23
- (b) A.fr. cont. : moy. = 0,04; σ = 0,08 (dont Somme, Pas-de-Calais)
- (c) Somme, Pas-de-Calais : moy. = 0,02; σ = 0,06

TABLE 5.3 – Taux de défaut de <e> (Tx.Ø) devant initiale vocalique en anglo-normand pour chaque période.

Période	Tx.Ø		Nb. textes
	__ #V	__ #C	
(1) 1100-1150	8%	1%	4
(2) 1151-1200	21%	5%	28
(3) 1201-1250	5%	0%	6
(4) 1251-1322	37%	7%	2

Or, sur les cartes par période (fig. 5.13, 5.15, 5.17 et 5.19), et en table 5.3 qui synthétise les résultats de l'anglo-normand, le taux moyen de défaut de <e> devant initiale vocalique en Angleterre semble suivre un dessin en dents de scie, passant de 8% à 21%, puis 5% avant de terminer à 37% (table 5.3). Le taux de défaut de <e> devant initiale consonantique varie moins, de 0 à 7%, mais son évolution est parallèle à celle des taux devant initiale vocalique.

Ces fortes variations du taux sont à mettre en relation avec le nombre de textes par période, dans la mesure où la baisse du nombre de textes semble filtrer arbitrairement les résultats, que ce soit en faveur (période 4) ou en défaveur (période 1 et 3) du taux de défaut de <e> devant initiale vocalique. En périodes 3 (1201-1250) et 4 (1251-1322), la variation de l'anglo-normand ne semble pas motivée par une évolution diachronique en tant que telle, mais plus par un effet de filtre dû au faible nombre de textes, les textes anglo-normands disposant d'une forte variabilité entre eux. Comme représenté en table 5.4, en période 3, quatre textes ont des taux de défaut de <e> devant initiale vocalique nuls, un texte a un faible taux (2%), mais un texte a un taux très supérieur à la moyenne (27%). C'est donc ce texte, *modw*, qui est responsable du caractère relativement élevé du taux de 5% d'occurrences en défaut de <e> devant initiale vocalique en anglo-normand dans la première moitié du XIII^e s.; sans lui, la moyenne de la région serait inférieure à la moyenne des régions sur l'ensemble du domaine.

TABLE 5.4 – Taux de défaut de <e> (Tx.Ø) devant initiale vocalique en anglo-normand de période 3 (1201-1250).

Texte	Tx.Ø	Compo.	Qualité
<i>aileg</i>	0%	1207	ms
<i>besant</i>	0%	1227	cr
<i>greg2</i>	0%	1215	ms1
<i>nicoa</i>	0%	1225	ms1
<i>best</i>	2%	1211	cr3
<i>modw</i>	27%	1230	cr

aileg : RAOUL DE HOUDENC (1207*). *Le roman des ailes*. Ms. G; Dees : Angleterre; compo. pic., ca 1207; ms. a.n., ca 1300; MS (RAOUL DE HOUDENC 1983).

besant : GUILLAUME LE CLERC DE NORMANDIE (1227). *Le besant de Dieu*. Compo. norm., 1227; ms. a.n., ca 1290; CR (GUILLAUME LE CLERC DE NORMANDIE 1973), DEAF : *BesantR*.

greg2 : JEAN LE DIACRE (1215). *La vie de Saint Gregoire le Grand*. Dees : [Angleterre]; compo. a.n., 1215; ms. a.n., ca 1215; CR (JEAN LE DIACRE 1883); DEAF : *AngVieGregM*.

nicoa : PSEUDO-NICODÈME (1225*). *L'Évangile de Nicodème*. Ms. A; Dees : [Angleterre]; compo. nil, ca 1225; ms. a.n., ca 1290; MS1 (PSEUDO-NICODÈME 1973); DEAF : *EvNicPraF*.

best : GUILLAUME LE CLERC DE NORMANDIE (1211*). *Bestiaire de la divine escripture*. Dees : Angleterre; compo. norm., ca 1211; ms. a.n., ca 1250; cr3 (GUILLAUME LE CLERC DE NORMANDIE 1892); DEAF : *BestGuillR*.

modw : ANON. (1230*). *Saint Modwenna*. Ms. O; Dees : Angleterre; compo. a.n., ca 1230; ms. a.n., ca 1250; CR (A. T. BAKER et BELL 1947); DEAF : *ModvB²*.

Au contraire, en période 2, si 11 textes anglo-normands disposent de taux de défaut de <e> devant initiale vocalique nuls, 17 textes disposent de taux allant de 9% à 100% (moy. de ces 17 textes : 35%). Aussi, le taux moyen de défaut de <e> devant initiale vocalique en anglo-normand en période 2, qui est de 16%, résulte de la moyenne entre 11 textes sans aucune occurrence défectueuse et de 17 textes dont le taux varie fortement. La situation de l'anglo-normand en période 2 est donc très hétérogène, l'écart-type s'établissant à 0,23, contre 0,08 en français continental à la même période, et 0,11 en anglo-normand en période 3. Il semble donc que les variations très marquées de l'anglo-normand soient dues au nombre de textes, celui-ci filtrant aléatoirement les textes comportant de nombreux défauts de <e> en période 3. Enfin, dans la dernière période, seuls deux textes anglo-normands sont présents, et leurs taux de défaut de <e> devant initiale vocalique sont très contrastés ($\sigma = 0,39$), comme indiqué en table 5.5.

TABLE 5.5 – Taux de défaut de <e> (Tx.Ø) devant initiale vocalique en anglo-normand en période 4 (1251-1322).

Texte	Tx.Ø	Compo.	Qualité
edmond	9%	1255	ms2
plainte	64%	1312	cr

Le pic observé dans la seconde moitié du xii^e s. tient donc en partie à la présence de très nombreux textes anglo-normands, lesquels sont plus susceptibles que les textes continentaux de disposer de valeurs extrêmes. Devant initiale consonantique, l'anglo-normand conserve une hétérogénéité supérieure à celle du domaine continental (5.65), mais la différence d'hétérogénéité est moindre.

(5.65) Écart-types devant initiale consonantique

- (a) A.n. : moy. = 0,04; $\sigma = 0,05$
- (b) A.fr. cont. : moy. = 0,01; $\sigma = 0,03$

Aussi, peu importe le contexte, l'anglo-normand est caractérisé non seulement par des taux moyens plus élevés, mais aussi par une dispersion des données plus large, et susceptible d'un important effet de filtre lorsque le nombre de textes décroît. Parallèlement, dans le contexte prévocalique, c'est aussi le dialecte qui dispose à la fois des taux les plus élevés et à la fois du nombre de textes à taux élevés le plus grand, se distinguant par-là nettement du domaine continental pris comme un ensemble plus cohérent ; cette différence est représentée en fig. 5.21, page suivante. La différence entre l'ensemble continental et l'ensemble anglo-normand est hautement significative ($p < 0,005$).⁴¹ Dans le contexte préconsonantique, tel qu'indiqué en fig. 5.22.a, l'anglo-normand ne se singularise qu'en termes relatifs : dans la seconde moitié du xii^e s., il dispose de plus de textes comportant des défauts de <e>, rapportés au nombre de textes sans défaut de <e>, que les dialectes continentaux, et cette concentration anglo-normande n'apparaît que dans cette seconde moitié du xii^e s.

41: Valeur p du test de Student en fig. 5.21.

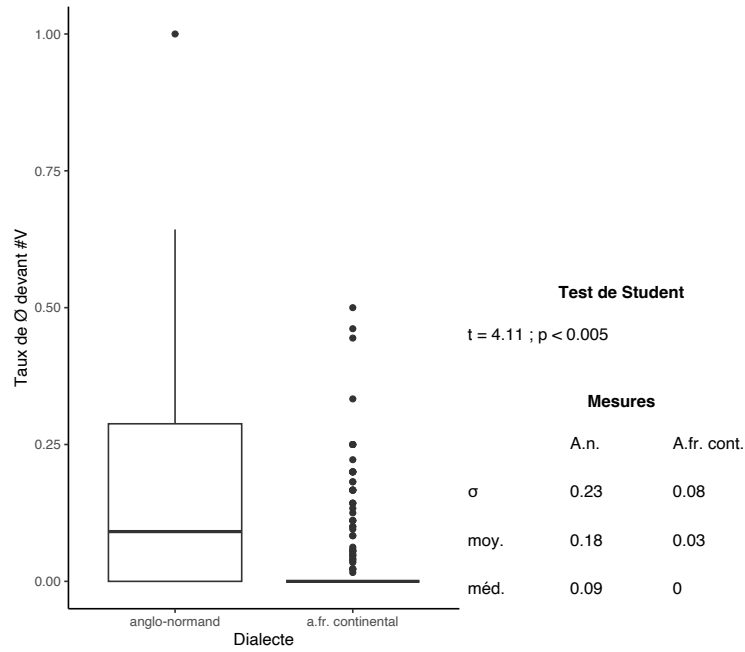


FIGURE 5.21 – Distinction entre anglo-normand et dialectes continentaux pour les taux de défaut de <e> des adjectifs féminins singuliers devant initiale vocalique.

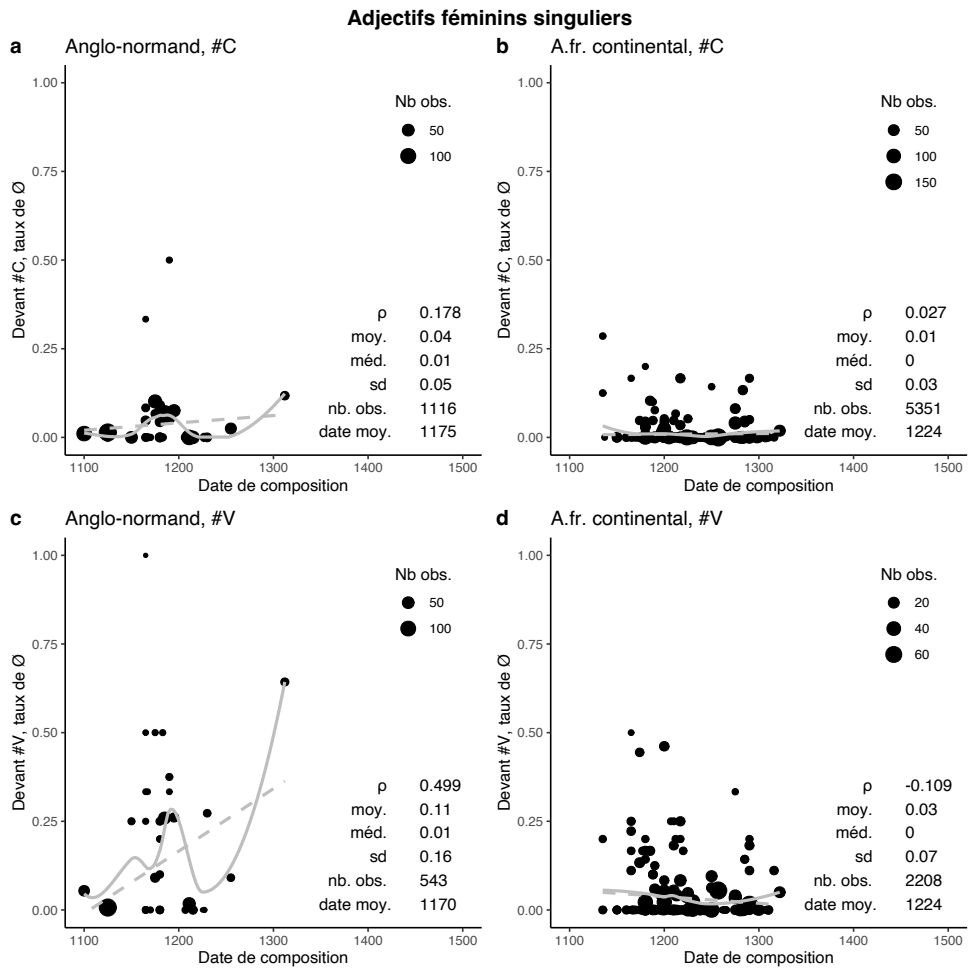


FIGURE 5.22 – Diachronie des taux de défaut de <e>, anglo-normand vs ancien français continental, adjectifs féminins singuliers.

Si l'anglo-normand est responsable d'une partie du pic des taux de défaut de <e> devant initiale vocalique dans la seconde partie du xii^e s., il n'est pas responsable de l'ensemble de cette concentration de taux élevés : on observe également de nombreux taux élevés en ancien français continental, tels que représentés en fig. 5.22.d. Dans cette période, en effet, trois autres régions ont des taux supérieurs à la moyenne régionale : la région Vendée, Deux-Sèvres (15%), la Normandie (11%), la Wallonie (10%) et la région Moselle, Meurthe-et-Moselle (8%) (fig. 5.15). En revanche, l'anglo-normand dispose de plus de textes comportant des défauts de <e> que de textes ne comportant pas de défaut de <e> pour cette période, et ces textes comportent plus d'occurrences de nos lemmes que les textes sans défaut de <e> (*nb.obs.* en fig. 5.22); singulièrement, l'anglo-normand n'a pas de texte sans défaut de <e> devant initiale vocalique entre 1180 et 1207. Ainsi, l'anglo-normand n'est pas seul responsable de la concentration de textes à forts taux de défaut de <e> devant initiale vocalique dans la seconde moitié du xii^e s., mais ce dialecte amplifie très fortement une tendance qui est beaucoup moins marquée dans les dialectes continentaux.

Comme mentionné *supra* à propos des taux par période, on observe en anglo-normand que les taux de défaut de <e> devant initiale consonantique suivent globalement l'évolution de ceux devant initiale vocalique, tout en étant bien inférieurs. Cela est également vrai en observant la diachronie non séparée en période, telle que présentée en fig. 5.22.a : un pic relatif s'observe dans la seconde partie du xii^e s., tout en étant bien moins élevé que le pic observable devant initiale vocalique. Sur le continent, une telle observation est également possible, mais les taux devant initiale consonantique sont si faibles (moy. = 1%) que la comparaison n'est pas très assurée.

Ainsi, devant initiale consonantique, l'anglo-normand est caractérisé par une forte variation interne, de 0 à 100% de défaut de <e> par texte ($\sigma = 0,23$ vs a.fr. continental $\sigma = 0,08$), peu homogène dans le temps, et visiblement en partie dépendante de la quantité de textes disponibles. Cette hétérogénéité se double de taux moyens globalement plus élevés que sur le continent. Cet élément, couplé au nombre important de textes disponible en période 2 (1151-1200 : 35 textes), est en bonne partie responsable du pic observé dans la seconde moitié du xii^e s. L'ensemble de ces observations peut être regroupé sous la forme des généralisations (5.66) et (5.67). En (5.66), les moyennes utilisées sont pondérées par le nombre d'occurrences des textes, comme en fig. 5.22; elles sont donc différentes de celles, non pondérées, utilisées en fig. 5.21.

(5.66) **ADJFÉM_{sg} : spécificité de l'a.n.**

L'anglo-normand (a) se distingue fortement des dialectes continentaux par (b) son hétérogénéité, (c) ses nombreux textes à forts taux de défaut de <e> et (d) son importante susceptibilité à un effet de filtre selon le nombre de textes disponibles.

- a) $_ \#V : t = 4,02 ; p < 0,005 ;$
- b) $_ \#V : \sigma = 0,23$ vs a.fr. continental $\sigma = 0,08 ;$
- c) $_ \#V : \text{moy.} = 11\%$ vs a.fr. continental 4%,
 $_ \#C : \text{moy.} = 4\%$ vs a.fr. continental 1% ;
- d) cf. (b), tables 5.4 et 5.5.

(5.67) **ADJFÉMSG : poids de l'a.n. dans les données.**

Étant donné (5.66) et le nombre élevé de textes anglo-normands (48), plus il y a de textes anglo-normands, plus les défauts de <e> sont nombreux.

5.3.6 Place de la variation au sein des ADJFÉMSG

Toutes les mesures citées *supra* concernent les défauts de <e> des occurrences des 45 lemmes susceptibles d'un défaut de <e> pour /ə/ final. Comme indiqué en table 5.1 *supra*, ces 45 lemmes ne représentent approximativement que 5% des lemmes d'adjectifs féminins singuliers du corpus,⁴² mais ces 5% de lemmes correspondent à 23% des occurrences d'adjectifs féminins singuliers, un certain nombre de ces 45 lemmes disposant d'une fréquence élevée. Nos 228 occurrences en défaut de <e> ne représentent que 0,6% de l'ensemble des occurrences d'adjectifs féminins singuliers.

La variation étudiée ici est donc doublement marginale, mais avec des degrés différents. Elle est fortement marginale au sein de l'ensemble des adjectifs féminins singuliers (0,6%), et peut en ce sens être assimilée à un simple bruit dans les données. Néanmoins, si l'on admet que seuls certains lemmes sont éligibles à cette variation, celle-ci, tout en demeurant marginale, n'est plus insignifiante au sein de cet ensemble (2%) — si nous ne pouvons définir de seuil de significativité pour ces valeurs dans l'absolu, leur distribution marginale prendra sens dans le cadre de l'analyse proposée dans la dernière partie de cette thèse. Si certains critères de cette sélection des lemmes vont de soi, comme le fait qu'il faut disposer d'un <e> de droit pour être qualifié pour le défaut de <e> (à l'exclusion, ainsi, des adjectifs épiciens oxytons *grant*, *fort*, etc., parfois très fréquents),⁴³ il n'en demeure pas moins que, au sein des lemmes *qui pourraient être en défaut de <e>*, un grand nombre d'entrées lexicales ne semblent pas concernées par cette variation.

Par ailleurs, l'existence des patrons diachroniques et diatopiques que nous avons identifiés (rôle de l'anglo-normand, fonctionnement en opposition centre - périphérie au début du xii^e s.), ainsi que l'effet marqué du contexte subséquent et l'inversion de cet effet dans le Sud-Est plaident en faveur d'une certaine logique, fût-elle imparfaite, dans la distribution des défauts de <e>. Aussi, la variation en question ne semble pas tenir uniquement à du *bruit dans les données* (au sens d'un faible pourcentage de données corrompues) mais à une *variation à bas bruit*. Si, face au faible niveau global de cette variation, il faut rester prudent dans les inductions que l'on peut proposer et dans le maniement des généralisations proposées *supra*, nous proposons néanmoins de considérer que cette variation a une certaine pertinence pour l'analyse grapho-phonologique.

(5.68) **ADJFÉMSG : poids de la variation.**

Rapportés à l'ensemble des ADJFÉMSG, les défauts de <e> sont extrêmement marginaux (a). Néanmoins, il semble qu'il s'agisse d'une *variation à bas bruit* plutôt que d'un simple *bruit de fond* (b).

- a) 0,6% des occ. d'ADJFÉMSG,
2,4% des occ. des 45 lemmes ;
- b) Cf. (5.61, 5.62, 5.63, 5.66, 5.67).

42: Le nombre total de lemmes du corpus est une estimation : il ne prend pas en compte les lemmes redondants, mais ne prend pas non plus en compte les nombreuses occurrences de lemmes rares que Tree Tagger n'a pas su identifier et auxquels il a attribué le lemme UNKNOWN.

43: *grant* ADJFÉMSG : 9137 occ. dans le NCA, soit presque autant que nos 45 lemmes à variation.

5.4 Conclusion

Dans ce chapitre, nous avons étudié la distribution des défauts de <e> des adjectifs féminins singuliers dans le corpus du NCA. Après avoir, dans l'introduction, justifié notre choix de travailler uniquement sur les adjectifs féminins singuliers (sec. 5.1.1), présenté les exclusions de données (sec. 5.1.2) et les formes relemmatisées (sec. 5.1.3), nous avons présenté les résultats de l'analyse que Comparalem nous a permis de mener (sec. 5.2). Après un bref point sur la typologie des lemmes et formes sujets à défauts de <e> (sec. 5.2.1), nous avons abordé les principales caractéristiques de la distribution de ces défauts de <e> : en diachronie (sec. 5.2.2), selon les lemmes et leur fréquence (sec. 5.2.3), selon le contexte phonologique (sec. 5.2.4), en diatopie (sec. 5.2.5), et à l'interface entre diachronie et diatopie (sec. 5.2.6). Enfin, nous avons procédé à une montée en généralisation, qui tient lieu de *discussion* des résultats avancés (sec. 5.3).

Cette généralisation nous a permis de réduire les faits observés à un ensemble de caractéristiques majeures. Au niveau des lemmes (sec. 5.3.1), nous n'avons pas noté de phénomène de diffusion lexicale observable dans nos données qui ne soit pas dû à la simple taille du corpus (5.57), ni de corrélation proprement linguistique entre la fréquence des lemmes et leurs taux de défaut de <e> (5.58) ; en revanche, pour trois lemmes relativement fréquents, l'inclusion de l'adjectif dans une structure *adj.* + *nom* relativement lexicalisée semble favoriser les défauts de <e>, du moins lorsque le second mot est à initiale vocalique (5.60). Dans ces collocations, l'application de l'élision semble témoigner de la réunion des deux mots au sein d'un même domaine computationnel. Ce même effet positif de la présence d'une voyelle à l'initiale du mot suivant se retrouve aussi indépendamment des collocations : les initiales vocaliques sont associées plus souvent à des défauts de <e> que les initiales consonantiques, et l'effet est statistiquement significatif (sec. 5.3.2). Cet effet n'est cependant pas stable dans le temps et dans l'espace, avec de fortes fluctuations de la puissance de l'effet selon les périodes et les régions, et un pôle Sud-Est dans lequel c'est le contexte préconsonantique qui est associé le plus souvent aux défauts de <e>. Du point de vue diachronique (sec. 5.3.3), le seul phénomène réellement observable est une concentration de taux positifs de défaut de <e> plus élevée dans la seconde moitié du XII^e s. que dans les autres périodes (5.62). Cet effet, s'il vaut *en général*, est renforcé par deux effets diatopiques : le poids des marges du domaine (sec. 5.3.4) et, singulièrement, le poids de l'anglo-normand (sec. 5.3.5).

Ainsi, alors que les marges comportent sur certaines périodes des taux de défaut de <e> plus élevés que ceux du reste du domaine, cela n'est pas vrai à toutes les périodes ; singulièrement, les marges génèrent beaucoup plus de défauts de <e> dans la seconde moitié du XII^e s. (5.63). Ce n'est pas non plus vrai indifféremment du contexte subséquent, cet effet des marges dans la seconde moitié du XII^e s. étant bien plus marqué devant initiale vocalique que consonantique. C'est donc l'expression de la singularité des marges, plus forte à cette période et particulièrement devant initiale vocalique, qui peut expliquer une partie du pic observé à cette époque ; le centre du domaine ne connaît pas une telle spécificité de la seconde moitié du XII^e s. Et, parmi les domaines marginaux, l'anglo-normand dispose d'un statut spécifique (5.66), et ce à au moins trois titres : par sa

forte variation interne, par ses taux de défaut de <e>, et par le nombre important de textes qu'il comporte. De fait, c'est bien l'anglo-normand qui a le plus fort effet sur ce pic de défaut de <e> dans la seconde moitié du XII^e s. (5.67), tandis que le poids de ce dialecte est quantitativement plus faible au XIII^e s.

Enfin, nous avons pris du recul pour ne plus simplement observer la variation au sein de nos 45 lemmes à variation, mais replacer cette variation dans l'ensemble des adjectifs féminins singuliers (sec. 5.3.6). De ce point de vue, la variation observée est fortement marginale ; néanmoins, elle ne semble pas tenir au hasard, dans la mesure où elle présente des distributions contextuelles, diatopiques et diachroniques (les secondes étant en partie dues aux premières) qui sont notables. Nous avons alors proposé de considérer cette variation non comme du *bruit dans les données*, mais comme une *variation à bas bruit* (5.68) : les défauts de <e> de ces 45 lemmes se jouent à une fréquence extrêmement basse, mais ne sont pas négligeables pour autant.

6.1 Introduction

Après avoir interrogé les défauts de <e> des noms et des adjectifs, nous nous intéressons à la catégorie des prépositions. Celle-ci montre en effet des taux de défaut de <e> (6% au sein de tous les lemmes, et 16% au sein des lemmes à variation) intermédiaires entre ceux de la catégorie des adjectifs féminins singuliers (0,6% au sein de tous les lemmes, et 2,43% au sein des lemmes à variation) et ceux des adverbes, qui fera l'objet du chapitre suivant.

La catégorie des prépositions dispose d'un statut syntaxique et lexical fort différent de celui des noms et adjectifs étudiés précédemment. Syntaxiquement, une préposition ne peut être le dernier élément de son constituant qu'en emploi absolu, emploi que nous n'avons pas rencontré dans les formes en défaut de <e> de notre corpus. Dans ses emplois canoniques, la préposition est donc *pré-posée*, ce qui induit, dans une langue dont l'accent frappe la droite des mots lexicaux et de certains mots fonctionnels, puis la droite des groupes de mots, que l'éventuel schwa qui terminerait une proposition ne serait pas nécessairement posttonique mais prétonique. Lexicalement, les prépositions sont des mots fonctionnels, mais dont la classe se développe fortement dans l'histoire du français, notamment en lien avec la raréfaction puis l'abandon de la déclinaison (FAGARD 2020, GGHF, p. 863-864). Ce développement, suivi d'un constant renouvellement de la catégorie indique des processus de grammaticalisation marqués, et concerne notamment les locutions prépositionnelles, dont il ne sera pas question ici puisque notre approche travaille sur la base de mots graphiques : dans, par exemple, *a cause d icelle* (« à cause de celle-ci », fierens, p. 53), le corpus annote une succession PRÉP-NOM-PRÉP-NOM et est aveugle au constituant grammaticalisé à valeur de préposition *a cause d(e)*. Par ailleurs, en ancien français, la relation entre la catégorie des prépositions et d'autres catégories, au premier rang desquelles celle des adverbes, est moins hermétique qu'en français moderne : de nombreux items comme *sor* et *desor* (fr.mod. « sor », « dessus ») peuvent fonctionner aussi bien comme adverbes que comme prépositions (PICOCHÉ et MARCHELLO-NIZIA 1994 [1989], p. 271; MÉNARD 1988 [1973], p. 262), alors qu'en français moderne, seuls *après*, *avant* et *devant* peuvent encore figurer dans les deux catégories (MARCHELLO-NIZIA 1999b, p. 120).

Enfin, étant en nombre restreint mais nécessaires à l'expression des relations paradigmatiques dans la syntaxe de moins en moins casuelle de l'ancien français, les prépositions sont caractérisées par des fréquences d'emploi élevées. Celles-ci sont peu homogènes : de nombreuses prépositions sont très rares tandis que quelques unes sont très fréquentes, parfois jusqu'à 1% des mots dans le corpus de la GGHF (FAGARD 2020, GGHF, p. 859). L'annotation du NCA fournit des résultats équivalents, avec 5 prépositions dont la fréquence absolue dépasse les 10 000 occurrences (cf. 6.1 *infra*). À titre de comparaison, dans le NCA, le lemme adjectival

FAGARD (2020). « Prépositions et locutions prépositives ».

fierens : ANON. (1489). *Cartulaire de l'hôpital Saint-Jacques de Tournai*. Dees : nil; compo. Tournai, ca 1489; ms. Tournai, ca 1489; MS (transcr. inédite non identifiée).

PICOCHÉ et MARCHELLO-NIZIA (1994 [1989]). *Histoire de la langue française*. MÉNARD (1988 [1973]). *Syntaxe de l'ancien français*

MARCHELLO-NIZIA (1999). *Le français en diachronie : douze siècles d'évolution*.

le plus commun est « grant » (13 000 occ.) et le lemme nominal le plus commun est « sire » (6013 occ.).¹ Le statut grammatical et les conditions d'emploi des prépositions se distinguent donc clairement de ceux des noms et adjectifs.

1: Nous excluons le lemme « pas », dont les occurrences annotées comme des noms sont plus nombreuses que celles de *sire*, mais qui comportent de très nombreuses erreurs d'annotation où *pas* est le forclusif.

6.1.1 Exclusions

Comme évoqué *supra*, la combinaison d'un nombre restreint de lemmes et d'une fonction syntaxique de plus en plus prégnante alors que la déclinaison s'affaïsse puis disparaît crée les conditions d'une fréquence d'usage élevée des prépositions, en général (en tant que catégorie), et en particulier pour certains lemmes prépositionnels. Ainsi, on trouve dans le NCA 204 lemmes de prépositions, correspondant à 256 856 occurrences. Sur ces 204 lemmes, les plus communs disposent de fréquences très élevées, et cinq disposent de fréquences supérieures à 10 000 occurrences (6.1). Suite aux exclusions contextuelles des requêtes qui nourrissent Comparalem, et aux deux relemmatisations appliquées (sec. 6.1.2), le nombre de lemmes sur lequel travaille Comparalem s'établit à 178, et le nombre d'occurrences à 230 885.

(6.1) Lemmes de PRÉP à plus de 10 000 occurrences.

- | | |
|-------------------------------------|---------------|
| (a) « de » | 88 230 occ. ; |
| (b) « a3_a » (fr.mod. « à ») | 52 835 occ. ; |
| (c) « en1_en » | 26 399 occ. ; |
| (d) « par » | 23 547 occ. ; |
| (e) « por2_por » (fr.mod. « pour ») | 21 327 occ. |

La catégorie des prépositions est donc caractérisée par un faible nombre de lemmes correspondant à un fort nombre d'occurrences. De ce fait, Comparalem n'a détecté que peu de lemmes à alternance <e>/Ø finale : il en a détecté 22. Parmi ces 22 lemmes, 8 correspondent en réalité à trois lemmes, Tree Tagger ayant multiplié les entrées et leurs combinaisons ; ils ont donc été relemmatisés sous trois entrées (sec. 6.1.2). Et, sur ces 17 lemmes, une minorité seulement correspond effectivement aux défauts de <e> que nous cherchons à étudier ici. Ainsi, nous avons exclu les lemmes où la variation observée ne correspond pas à l'absence d'un <e> qui serait de droit, étant donnée l'étymologie, pour une préposition. Les catégories de ces exclusions sont données en (6.2) et rapidement développées *infra*.

(6.2) Causes d'exclusions et exemples

(a) Composition sur second terme adjectival

— « atout » : AD+TOTUS, -A → *atout*, -e

de bel aage atot le sanc

et atoute lor force fussent

« ils auraient été dans la force de l'âge, avec toute la vivacité du sang »

(chret1, p. 63, trad. POIRION : CHRÉTIEN DE TROYES 1994, p. 729)

chret1 : CHRÉTIEN DE TROYES (1180*). *Le Chevalier au lion & Perceval*. Dees : Hainaut ; compo. champ. mérid., ca 1180 ; ms. champ., ca 1238 ; MS (transcr. inédite R. L. H. Lops) ; contenu présent ailleurs dans le corpus : incipit du *Chevalier au lion* (v. 1-998) dans yvf, et *Perceval* (v. 69-2000) dans perr.

- « enmi » : IN+MEDIUS → *enmi* mais qq. occ. *enmie*
s est pamee et giest morte et descouloree enmie le lit morte
sanz vie
 « [Elle] est tombée en pamoison et git morte, pâle, au
 milieu du lit, comme morte et sans vie »
 (vergif, p. 18, trad. pers.)
- « parmi » : PER+MEDIUS → *parmi* mais qq. occ. *parmie*
ainz l enporte grant aleure parmie la grant forest
 « Ainsi [le cheval] l’emporte à vive allure dans la
 grande forêt »
 (perf, p. 6, trad. pers.)
- (b) <a> final pour /ə/
 — « jost1 » : type *josta* au Sud-Ouest
iezent iosta l outer
 « ils étaient à côté de l’autel »
 (poit, p. 75, trad. pers.)
- (c) <e> final pour /e/
 — « parmi » : qq. occurrences sans <i>
fu li vasaus perme la sale
 « Le vassal traversa la salle »
 (alia, p. 4, trad. pers.)
- (d) Erreur d’annotation en POS
 — « en2_ne » : ADV ou CONJ
ainc nus ne t eut en sa memoire n ainc nus merci ne te cria
 « Personne ne se souvient de toi, personne ne te
 demanda merci », ou
 « Personne ni ne se souvient de toi ni ne te demanda
 merci »
 (mir, p. 21, trad. pers.)
- (e) Erreur d’annotation en lemmes
 — « avant » : 1 occ. correspond à « avec »
il vivoit ave sa femme en une mesme maison
 « il vivait avec sa femme dans la même maison »
 (abe, p. 55, trad. pers.)
 — « avers2 » : 1 occ. correspond à « aversque »
de dangier neant ne monte aversque de ma fille
 « tu ne montes pas sans danger auprès de ma fille »
 (rose, p. 50, trad. pers.)

En (6.2 a), les lemmes *atout*, *enmi* et *parmi* ne sont pas censés disposer d’un <e> pour schwa final ; ils sont tous composés par la fusion d’une préposition (*a*, *en*, *par*) et d’un mot lexical (*tout*, *mi*). Le second terme de la première composition, *tout*, est censé être oxyton s’il est au masculin ou n’est pas flechi (TOTUM → *tout*), mais est paroxyton s’il est féminin (TOTA → *toute*) ; le second terme des deux autres compositions est dans la même situation : de l’adjectif latin MÉDIUS, -A, -UM dérive l’adjectif a.fr. *mi*, -e (« qui se trouve au milieu », DMF). L’alternance entre formes bisyllabiques et trisyllabique *atout* ~ *atoute*, *enmi* ~ *enmie* et *parmi* ~ *parmie* ne tient donc pas à un défaut de <e>, et l’origine précise de ce <e> final ne nous concerne pas. De plus, graphies médiévales et éditions fluctuent quant au découpage de ces unités. Sur le vers cité pour « atout » en (6.2 a), POIRION (CHRÉTIEN DE TROYES 1994, p. 729), qui édite un autre ms. de base,

vergif : ANON. (1240*). *La chastelaine de Vergi*. Ms. F ; Dees : Normandie ; compo. norm., ca 1240 ; ms. pic., ca 1318 ; MS1 (STUIP 1970) ; DEAF : ChastVergis.

perf : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS F ; Dees : nil ; compo. champ. mérid., ca 1180 ; ms. Est, ca 1250 ; MS1 (transcr. inédite R. L. H. Lops).

poit : ANON. (1217*). *Chronique traduite de l’Historia Karoli Magni et Rotholandi*. Dees : Vendée, Deux-Sèvres ; compo. saint., ca 1217 ; ms. saint., ca 1238 ; MS1 (AURACHER 1877) ; DEAF : Turpin¹A.

alia : ANON. (1185*). *Roman d’Alexandre*. Ms. A ; Dees : Vendée, Deux-Sèvres ; compo. pic. et occ. orient., ca 1185 ; ms. poit., ca 1225 ; MS (ANON. 1937) ; DEAF : AlexArsL.

mir : GAUTIER DE COINCI (1224*). *Les miracles de Notre Dame de Soissons*. Ms. R ; Dees : Aisne ; compo. Soissonnais, ca 1224 ; ms. pic. mérid., ca 1310 ; CR (GAUTIER DE COINCI 1966) ; DEAF : CoincyI10K.

abe : JEAN DE MEUN (1288*). *Traduction de la première épître de P. Abélard*. Dees : région parisienne ; compo. frc., ca 1280 ; ms. Paris, ca 1395 ; MS3 (JEAN DE MEUN 1934).

rose : GUILLAUME DE LORRIS (1230*). *Le roman de la rose*. Dees : Nièvre, Allier ; compo. traits orl., ca 1230 ; ms. orl., ca 1285 ; CR2 (GUILLAUME DE LORRIS 1934) ; DEAF : Rose_LLec.

CHRÉTIEN DE TROYES (1994). *Œuvres complètes*.

le ms. P (BFN fr. 794), obtient *a tot* et *a tote*, ce qui dissocie la préposition du second terme, et permet à celui-ci de s'accorder en tant qu'adjectif. Le ms. édité par Poirion est le même que celui de chret2 et de PercevalKu (dans le corpus de la BFM), et les trois éditeurs obtiennent, sur le ms. P, la même lecture *a tote*. C'est donc la fusion des mots dans les éditions du ms. de chret1, P8 (BFN fr. 1450), qui crée l'alternance <e>/∅ de la préposition composée *atout*, ce qui rend l'occurrence d'autant plus complexe à analyser. Au demeurant, ces formes sont toujours marginales dans notre corpus (*atoute* : 6% des occurrences du lemme, *enmie* et *parmie* : 2% des occurrences des lemmes) et ne semblent donc pas témoigner d'une construction régulière depuis la variante fléchie de l'adjectif.

Comme pour les noms et les adjectifs, deux motifs d'exclusion tiennent au rapport entre graphèmes et phonèmes : on trouve dans le Sud-Ouest des graphies <a> au lieu de <e> (*iosta* au lieu de *juste*, *contra* au lieu de *contre*, *dejosta* au lieu de *dejoste*, *encontra* au lieu de *encontre*, etc. ; 6.2 b), que Comparalem analyse comme un défaut de <e> (il cherche un <e> qu'il ne trouve pas), alors qu'il s'agit plus probablement d'un système graphique dans lequel les finales atones peuvent être notées <a>. On trouve également quelques cas où <e> final note la voyelle /e/, comme *parme* pour *parmi* (6.2 c), *apre* pour *aprez*, etc.), sans que Comparalem n'ait les moyens de comprendre qu'il s'agit là d'une voyelle différente du schwa noté <e> qu'il cherche.

Ces résultats comportent également des erreurs d'annotation, que ce soit en parties du discours ou en lemmes ; par exemple, l'occurrence *ne* en (6.2 d), adverbiale ou conjonctive, est annotée comme étant la préposition (erreur de POS) *en* (erreur de lemme). D'autres erreurs ne concernent que le lemme, avec par exemple *ave* lemmatisé comme étant une forme de « avec », ou *aversque* lemmatisé comme étant une forme de « avers » (6.2 e). Dans ce dernier cas, si les deux lemmes « avers » et « aversque » sont bien liés, ils ne sauraient être considérés comme constitués du même matériel phonologique.

Enfin, dans certains cas, c'est l'édition du NCA qui provoque l'exclusion d'une ou plusieurs formes. Ainsi, toutes les formes précédant immédiatement un chiffre sont exclues *via* le fichier d'exclusion des occurrences, dans la mesure où certains textes du NCA remplacent les chiffres romains par des successions de lettres, souvent *yy* ; sans cette exclusion, Comparalem considérerait que ces mots sont suivis d'une initiale vocalique (<y>). Dans quelques cas, l'édition est fautive non pas par normalisation mais par lacune ; de tels cas n'ont pu être découverts que lors d'investigations approfondies et nous ne pouvons être sûrs de les avoir tous éliminés. Un exemple paroxystique de cela peut être trouvé au vers 203 de *Roland* : alors que les éditeurs s'accordent plus ou moins sur la leçon à établir, le ms. présente un étrange cas de fusion entre deux mots : *païen(s)* et *enveiat* partagent la même séquence <en> : *païenveiat* (6.3 et fig. 6.1). Ici, le texte du NCA, contrairement à celui des autres éditions que nous avons consultées, non seulement ne restitue pas un doublement de <en>, mais supprime le mot *païen(s)* ainsi que le possessif qui le précède — rien dans le fichier XML n'indique la lacune. Dans ce cas précis, le NCA indique donc un *de* placé devant initiale vocalique (*de enveiat*), alors que ce *de* est devant initiale consonantique (*de ses*) sur le ms. L'édition du NCA reposant sur celle de Segre, il semble probable que Dees ait voulu éviter

chret2 : CHRÉTIEN DE TROYES (1180*). *Le Chevalier au lion & Perceval*. Ms. H (*Le Chevalier au lion*) & Ms. A (*Perceval*) ; Dees : Haute-Marne ; compo. champ. mérid., ca 1180 ; ms. champ., ca 1235 ; MS (transcr. inédite R. L. H. Lops & CHRÉTIEN DE TROYES 1975) ; contenu présent ailleurs dans le corpus : incipit du *Chevalier au lion* (v. 1-998) dans *yvh*, et *Perceval* (v. 1-2000) dans *pera* ; DEAF (*Perceval*) : PercL.

PercevalKu : CHRÉTIEN DE TROYES (1183*). *Conte du Graal (Perceval)*. Compo. ca 1181-1185, ms. ca 1235, dial. champ. (CHRÉTIEN DE TROYES 2009a).

d'intégrer l'interpolation que Segre signale, mais que ce faisant il ait également supprimé le mot précédant cette interpolation.

- (6.3) NCA : *de enveiat quinze*
 BFM : *De ses païens enveiat quinze* (id. BÉDIER 1947)
 SHORT : *de ses païens vos en enveiat quinze*
 SEGRE : *De ses païen<s vos en en>veiat quinze*
 Dipl. : *D efes païenveiat quinze*
 « il envoya quinze de ses païens »
 (Roland : rolandox (NCA), roland (BFM), SHORT 1990 et SEGRE 1971, v. 203, trad. SHORT 1990)

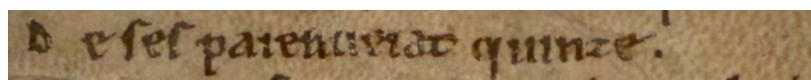


FIGURE 6.1 – *Roland*, ms. d'Oxford, fol. 4v, l. 5 (v. 203). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

Sur les 22 lemmes détectés par Comparalem, nous avons exclu 16 entrées, qui correspondent toutes à l'un des types en (6.2), et avons exclu ponctuellement 85 occurrences ; celles-ci correspondent à l'un des types en (6.2) mais ne représentent pas les seules occurrences en défaut de <e> de leur lemme.

6.1.2 Relemmatisation

Nous avons procédé à la relemmatisation de trois séries de lemmes, pour lesquels Tree Tagger propose plusieurs étiquettes pour ce qui correspond, en réalité, à un seul lemme linguistique. Ces lemmes sont, au niveau linguistique, « encontre » (*en+contre*),² « entre » et « sor » (fr.mod. « sor »), qui sont ici chacun divisés en plusieurs étiquettes : *encont re1* et *encontrer* ; *entre2_entre*, *entre2* et *entrer* ; *sor1_sor*, *sor2* et *sor2_sur1*. Une fois les occurrences en défaut de <e> vérifiées manuellement et la communauté du lemme, au sens linguistique, établie, nous avons donc fusionné ces couples de lemmes en une seule entrée par couple, comme indiqué en (6.4).

- (6.4) Tree Tagger Nouveau lemme
encont re1, *encontrer* → *encont re_ relemma*
entre2_entre, *entre2*, *entrer* → *entre_ relemma*
sor1_sor, *sor2*, *sor2_sur1* → *sor_ relemma*

Le cas de « entre » est instructif sur le comportement de Tree Tagger : sur les 461 occurrences prépositionnelles de l'étiquette *entre2*, 448 correspondent à *entr* ou *antr*, les autres formes étant dotées d'un -s (*entres*, *antres*) ou étant des erreurs de lemmatisation. Au contraire, le lemme *entre2_entre* ne contient que des formes *entre* et *antre* (1510 occ.) et le lemme *ent rer* comprend 29 formes *entre* et *antre*, 26 formes *entra* (avec <a> en finale atone dans le Sud-Ouest du domaine : *alia* et *poit*) et une forme mal annotée (*entrer*, INF). Parallèlement, une seule occurrence du corpus est annotée *ent re*, et ce n'est ni une préposition ni le mot *entre*, mais la conjonction complexe *entretantque* (stsilv, p. 21). Tree Tagger a donc été incapable de repérer la communauté de lemme entre les formes

rolandox : ANON. (1100*). *La Chanson de Roland*. Ms. O ; Dees : Angleterre ; compo. N.-O., ca 1100 ; ms. a.n., ca 1137 ; CR1 (SEGRE 1971) ; DEAF : Ro1S.

roland : Gérard MOIGNET, éd. (1972b). *Chanson de Roland*. Paris : Bordas. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 24-2-2014. URL : <http://catalog.bfm-corpus.org/roland>.

2: DMF : « Au contact de, en face de, en direction de, en opposition à » ; GD : « contre, envers » ; AND : « against, in conflict with ; in opposition to ; in breach of ; at the expense of », etc.

alia : ANON. (1185*). *Roman d'Alexandre*. Ms. A ; Dees : Vendée, Deux-Sèvres ; compo. pic. et occ. orient., ca 1185 ; ms. poit., ca 1225 ; MS (ANON. 1937) ; DEAF : AlexArsL.

poit : ANON. (1217*). *Chronique traduite de l'Historia Karoli Magni et Rotholandi*. Dees : Vendée, Deux-Sèvres ; compo. saint., ca 1217 ; ms. saint., ca 1238 ; MS1 (AURACHER 1877) ; DEAF : Turpin¹A.

en défaut de <e> et les formes où le <e> est bien présent, et il a fallu que nous la rétablissions nous-même après avoir exploré les formes, indépendamment des lemmes.

Enfin, pour « sor », si cette opération nous a semblé nécessaire dans la mesure où ces trois étiquettes contenaient de nombreuses occurrences de la préposition « sor », elle demeure problématique. Comme l'indique le caractère composé des étiquettes *sor1_sor* et *sor2_sur1*, ces étiquettes témoignent d'une difficulté du lemmatiseur à établir le lemme des occurrences. Or, comme nous le développerons plus bas, non seulement « sor » n'est pas toujours bien annoté en parties du discours, avec une séparation parfois discutable entre occurrences prépositionnelles et adverbiales, mais l'étiquette *sor2_sur1* recoupe aussi de nombreuses erreurs de lemmatisation (p.ex. *sor2_sur1* pour « sœur »). Si, le plus souvent, la combinaison du lemme et de la partie du discours permettent d'éliminer ces formes (*sor2_sur1* pour « sœur » est annoté comme un nom), il n'est pas assuré que ce soit toujours le cas, et le nombre trop élevé d'occurrences ne nous a pas permis de conduire une vérification manuelle de chaque occurrence.

6.2 Présentation des résultats

Les six lemmes retenus représentent plus de 3% des 175 étiquettes de lemmes associées à des formes prépositionnelles, comme indiqué en table 6.1. Si ce nombre de 175 lemmes correspondant à des prépositions est probablement trop élevé pour une catégorie fonctionnelle en inventaire restreint, nous avons déjà expliqué (sec. 4.2 et 5.2) qu'il n'avait pas fait l'objet d'une vérification de notre part et que Tree Tagger avait tendance à multiplier les combinaisons d'étiquettes en cas d'hésitation. FAGARD (2020, GGHF, p. 866) ne mentionne qu'une cinquantaine de prépositions en ancien français (contre une centaine sur l'ensemble du corpus de la GGHF, *id.* p. 859), ce qui vient confirmer la surestimation du nombre de lemmes de Tree Tagger. Au total, ce sont 230 885 occurrences qui sont annotées comme étant des prépositions dans le corpus. Nos six lemmes correspondent à 83 852 occurrences, soit 36% de l'ensemble des occurrences prépositionnelles. Le lemme « de » est à lui seul responsable de cette fréquence très élevée des lemmes en défaut de <e>, avec 77 369 occurrences présentes dans les contextes interrogés (92% des occurrences de nos six lemmes à variation). Sur ces six lemmes que nous avons retenus comme démontrant une variation par défaut de <e>, 13 830 occurrences sont en défaut de <e>. Ces occurrences en défaut de <e>

TABLE 6.1 – Quantité de prépositions, de lemmes à variation, et d'occurrences en défaut de <e>.

	Nb.	Nb.var.	Taux	Nb.Var.Ø	Taux.Ø/Nb.
Lemmes	175	6	3,43%		
Occ.	230 885	83 852	36,32%	13 830	5,99%
<i>Tx_Ø</i> (Nb.Var.Ø/Nb.var) :					16,45%

représentent 6% de toutes les occurrences prépositionnelles, et plus de 16% des occurrences des six lemmes en question. La variation est donc d'une ampleur bien plus importante que celle observée sur les adjectifs, où les formes en défaut de <e> ne représentaient que 0,6% des occurrences adjectivales au féminin singulier et 2,4% des occurrences des lemmes retenus.

6.2.1 Typologie

Les six lemmes retenus pour leurs occurrences en défaut de <e> sont « contre », « de », « desor » (fr.mod. *dessus*), « encontre » (sous la forme *encontre_relemma*, cf. sec. 6.1.2), « entre » et « sor » (sous la forme *sor_relemma*); deux lemmes sur six sont donc composés par la préfixation d'une autre préposition (*desor* : *de* + *sor*; *encontre* : *en* + *contre*). Ces prépositions disposent de caractéristiques diverses. *De* est monosyllabique, et le défaut de <e> la réduit donc à sa seule attaque. « Sor » devrait disposer d'un <e> de droit, étant formé depuis un étymon SUPER comportant un groupe /pr/ suite à la chute de la voyelle posttonique, ou depuis un étymon SUPRA comportant, en plus du groupe /pr/, un -A final.³ Néanmoins, si les formes bisyllabiques sont bien les seules présentes dans les premiers textes (*soure* dans *Eulalie*, v. 12; *sore* dans *Jonas*, l. 145), elles sont en concurrence avec des formes monosyllabiques (type *sor*) dès *Roland*, et ces formes sans <e> entraîneront une ré-analyse de la forme sous-jacente pour aboutir au *sur* du fr.mod. Comme en témoigne la forme des étiquettes du NCA correspondant à « sor » (*sor1_sor*, *sor2*, *sor2_sur1*), la disparition du <e> pour ce lemme est déjà majoritaire en ancien français (cf. § *Sor* p. 194; table 6.2; sec. 6.2.4.2) Il en va de même pour « desor », qui apparaît sans <e> final dès la fin du XII^e s. dans notre corpus.

Phonologiquement, les prépositions « contre », « encontre » et « entre » disposent de segments posttoniques contraints, dans la mesure où elles se terminent, encore en a.fr., par un groupe *obstruante* + *liquide*. Dans ce cas, le défaut de <e> laisse le groupe OL en position finale, ce qui donne 13 occurrences *contr*, 7 occurrences *encontr* et 435 occurrences *entr*, toutes situées devant initiale vocalique. De manière analogue, la réduction de la préposition *de* à une simple attaque consonantique impose des conditions phonotactiques, et les défauts de <e> de *de* n'apparaissent presque systématiquement que devant initiale vocalique. Les autres prépositions, c'est-à-dire « sor » et « desor », se terminent par une séquence Cə où C est par ailleurs licite en position finale absolue : on ne compte pas les mots qui, en ancien français, se terminent par un simple /r/ (ne serait-ce qu'au sein des prépositions : *par*, *pur*, etc.). Par ailleurs, et contrairement à ce que nous avons pu observer pour les adjectifs féminins singuliers, ce /r/ final n'est pas sujet à des alternances de voisement en position finale, la rhotique médiévale étant *a priori* insensible à la question du voisement, étant apico-alvéolaire. Enfin, puisqu'il s'agit ici de prépositions, il n'est pas question d'aménagements phonotactiques liés à l'adjonction d'un morphème flexionnel, comme cela peut être le cas pour les noms et les adjectifs (p.ex. avec *corp+s* → *cors*, cf. sec. 10, p. 138), puisque les prépositions ne sont pas sujettes à l'ajout de tels morphèmes.

3: On rencontre parfois l'idée que les formes *sor* viendraient d'un étymon SUPER tandis que les formes *sore* seraient issues d'un étymon SUPRA (MÉNARD 1988 [1973], p. 290), mais le FEW indique que c'est SUPER qui est l'étymon en Gaule. La distinction est au demeurant non pertinente pour nous, puisque dans les deux cas le <e> est de droit : SUPER, comme PATER → *pere*, SEMPER → *sempre* OU PAUPER → *povre*, *porre*, comporte un groupe *obstruante* + *liquide* en position finale une fois la voyelle E posttonique amuïe.

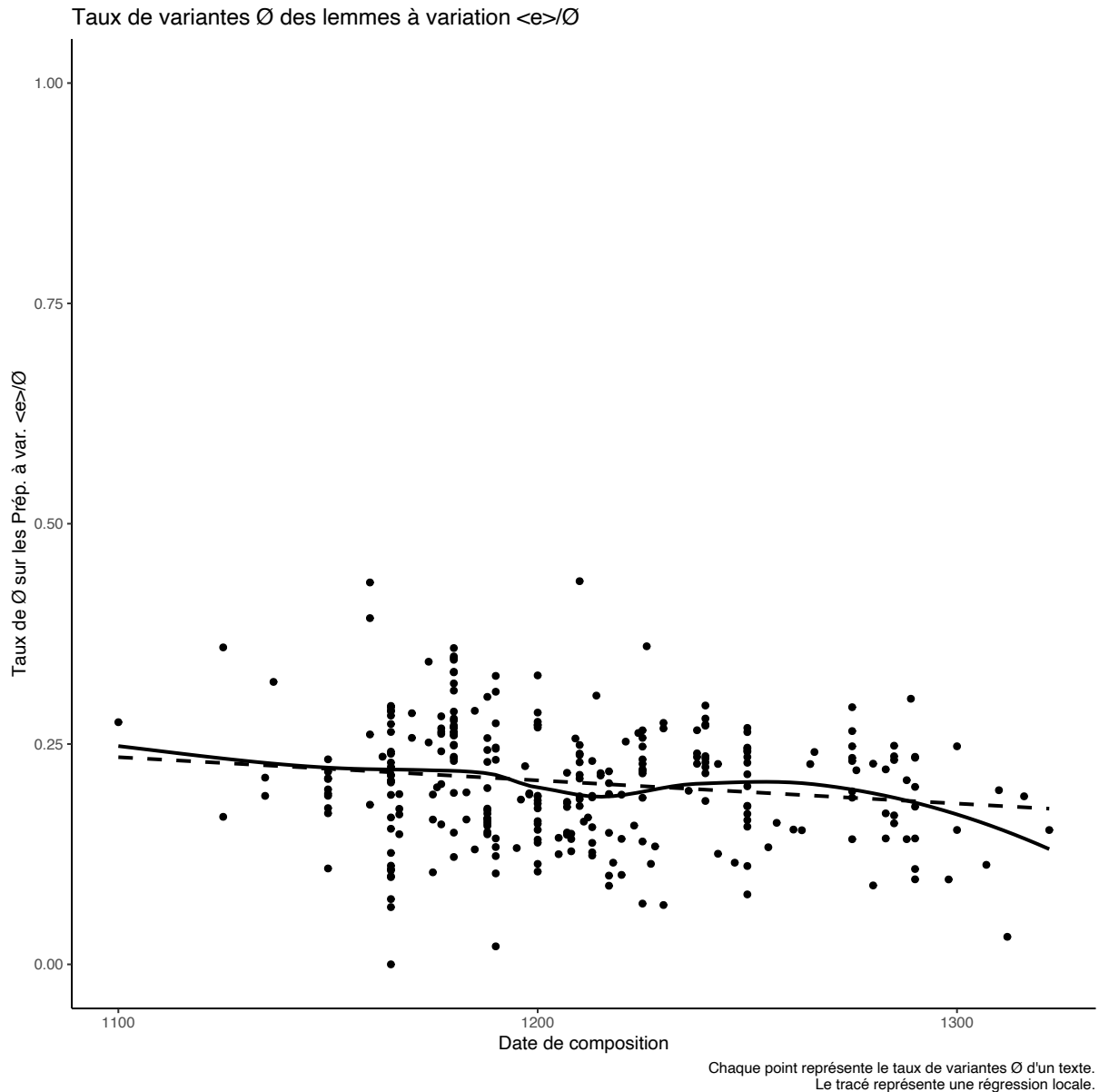


FIGURE 6.2 – Distribution diachronique des variantes de prépositions en défaut de <e>.

6.2.2 Distribution diachronique générale

La distribution diachronique générale des défauts de <e> sur les prépositions est peu marquée, comme représentée en fig. 6.2. En laissant de côté les autres variables (contexte, diatopie, fréquence), le taux moyen de défaut de <e> par texte est relativement stable, avec une moyenne sur l'ensemble de la période de 20%, un minimum à 0% (1 texte) et un maximum à 43% (1 texte). Ce taux moyen de 20% cache en réalité une situation très différente selon les lemmes (sec. 6.2.3) et selon que le mot suivant commence par une initiale vocalique ou consonantique (sec. 6.2.4 *infra*), avec un effet diachronique spécialement marqué en anglo-normand (sec. 6.2.5 *infra*).

6.2.3 Distribution des lemmes en défaut de <e> : diachronie et fréquence

Les graphiques en fig. 6.3 et 6.4 en pages suivantes représentent, pour chaque texte, le taux de défaut de <e> de chaque lemme, devant initiale vocalique en fig. 6.3, et devant initiale consonantique en fig. 6.4. Ces graphiques évoquent un effet contextuel très fort, qui sera confirmé *infra* (sec. 6.2.4), avec beaucoup plus de défauts devant initiale vocalique que consonantique. Relativement à cet effet, deux classes de lemmes se distinguent, avec d'une part les lemmes qui répondent très fortement à cet effet, et d'autre part ceux qui semblent indifférents au contexte.

Le lemme sur lequel le contexte subséquent a le plus de force est « de » : alors qu'il ne connaît presque aucun défaut de <e> devant initiale consonantique (fig. 6.4 : 31 défauts de <e> sur 66 087 occurrences), ses défauts de <e> sont très largement dominants devant initiale vocalique (fig. 6.3 : 9856 défauts sur 11 288 occ., soit 87%). C'est également le cas de « entre », qui ne connaît aucune forme en défaut de <e> devant initiale consonantique, et dont les deux tiers des formes devant initiale vocalique sont en défaut de <e> (435 formes du type *entr* contre 204 formes du type *entre* __#V); le groupe *obstruante* + *liquide* posttonique n'a donc ici aucun aspect protecteur vis-à-vis de l'élision. Dans une moindre mesure, cela concerne aussi « contre » et « rencontre », qui ne connaissent pas de défaut devant initiale consonantique mais dont les défauts devant initiale vocalique sont minoritaires (« contre » : 13 défauts sur 106 occurrences prévocales, soit 12%; « rencontre » : 7 défauts sur 20 occurrences prévocales, soit 35%).

En revanche, les lemmes « sor » et « desor » semblent insensibles à cet effet du contexte prévocale comme favorisant les défauts de <e> : « sor » connaît autant de défauts de <e> devant initiale vocalique (96%) que consonantique (97%), et « desor » n'apparaît que devant initiale consonantique, avec 7 occurrences en défaut de <e> contre une occurrence non défectueuse. Aussi, le lemme « sor » et le lemme « desor » (au demeurant formé sur « sor ») ne se comportent pas de la même manière que les lemmes « de », « entre », « contre » et « rencontre ».

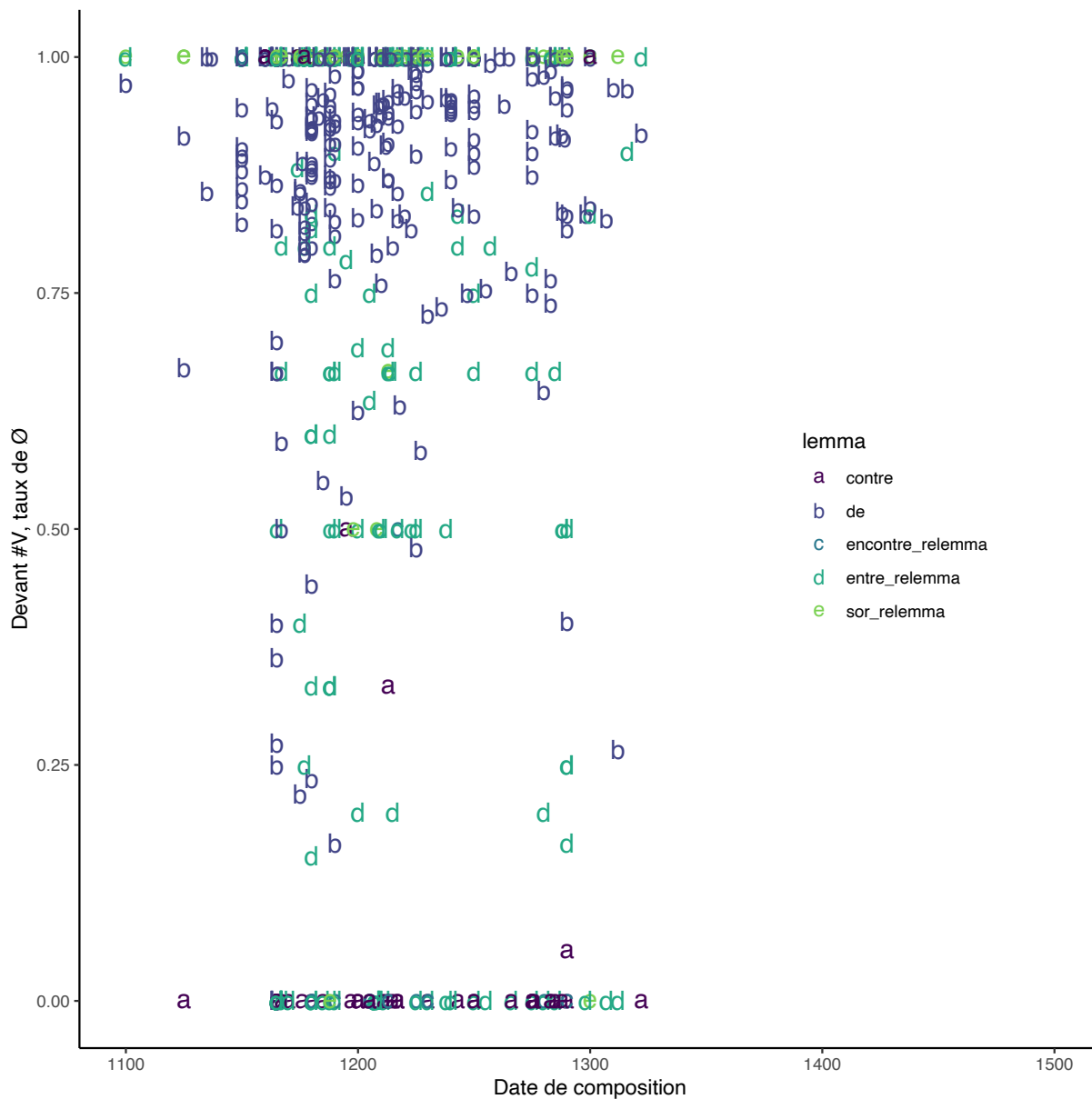
6.2.3.1 Lemmes attestés en défaut de <e> dès *Roland*

En diachronie, les premières occurrences de prépositions en défaut de <e> apparaissent pour les lemmes « de », « entre » et « sor » dès *Roland*.

D (« de »). Les attestations de *de* en défaut de <e> ne se trouvent, dans *Roland*, que devant initiale vocalique, comme par exemple en (6.5).

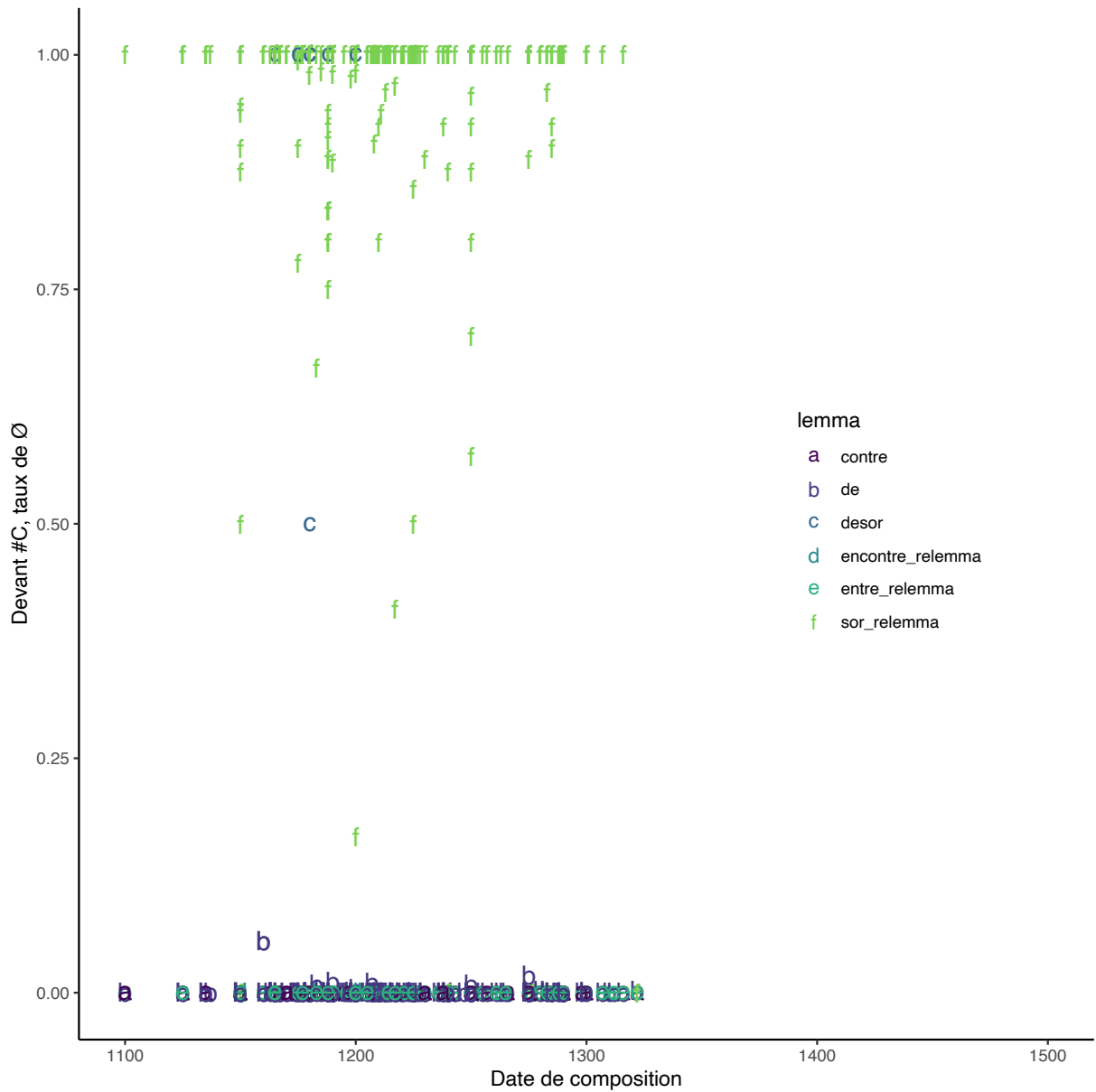
(6.5) *d or e d argent [quatre cenx] muls chargez*
« quatre cents mulets chargés d'or et d'argent »
(rolandox, v. 31, trad. SHORT 1990)

Si, dans *Roland*, le taux de défaut de <e> devant initiale vocalique est élevé (98%), le défaut n'est pas absolu : trois occurrences de « de » devant initiale vocalique comportent leur voyelle. Ces trois occurrences ont, chacune, un statut différent. La première occurrence (6.6) apparaît au



lemma	nb.e	nb.∅	tot.	tx.∅
sor_relemma	6	152	158	0.96
de	1435	9856	11291	0.87
entre_relemma	204	435	639	0.68
encontre_relemma	13	7	20	0.35
contre	93	13	106	0.12

FIGURE 6.3 – Distribution diachronique des alternances <e>/∅ selon les lemmes, prépositions suivies d’une initiale vocalique.



lemma	nb.e	nb.Ø	tot.	tx.Ø
sor_relemma	94	3338	3432	0.97
desor	1	7	8	0.88
de	66056	22	66078	0.00
entre_relemma	1215	0	1215	0.00
contre	899	0	899	0.00
encontre_relemma	6	0	6	0.00

FIGURE 6.4 – Distribution diachronique des alternances <e>/Ø selon les lemmes, prépositions suivies d’une initiale consonantique.

v. 149; dans l'édition du NCA, le premier mot du vers (*pa* pour *par*) est manquant,⁴ mais le ms. porte bien une version non éliée de la préposition : *de ocire*, comme reproduit en fig. 6.5.⁵ Dans l'édition utilisée par la BFM (reposant sur MOIGNET 1972a), le premier mot est bien présent, mais l'éliision de *de* est introduite alors qu'elle ne figure pas sur le ms.; dans leurs éditions, BÉDIER (1947, p. 14) et SHORT (2007, p. 36) font de même. Évidemment, l'édition de SEGRE (1971, p. 27) sur laquelle se fonde le NCA comporte bien le mot initial *Pa<r>*, ainsi que l'absence d'éliision : *de ocire*. Sur cette occurrence, le <e> est donc bien présent sur le ms., et certains éditeurs appliquent d'autorité une éliision à son endroit, le remplaçant par une apostrophe. En effet, si le <e> est conservé, le vers devient hypermétrique, avec 11 syllabes métriques au lieu de 10.

- (6.6) NCA : *num de ocire i metrai un mien filz*
 BFM : *Pa num d'ocire i metrai un mien filz*
 SEGRE : *Pa<r> num de ocire i metrai un mien filz;*
 SHORT : *Par num d'ocire i metrai un mien filz,*
 Dipl. : *p a nū· de ocire i metrai un mien filz*
 « j'y enverrai un de mes propres fils, au péril de sa vie⁶; »
 (*Roland* : rolandox (NCA), roland (BFM), SEGRE 1971, p. 27, SHORT 1990, p. 36, v. 148, trad. SHORT 1990)

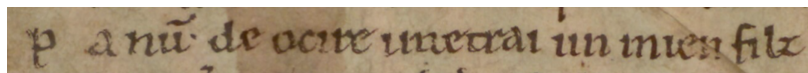


FIGURE 6.5 – *Roland*, ms. d'Oxford, fol. 3v, l. 8 (v. 148). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

La deuxième occurrence où <de> est noté devant voyelle est reproduite en (6.7). Elle appelle également une éliision pour que le vers ne soit pas hypermétrique; cette éliision est appliquée, toujours en remplaçant le <e> du ms. par une apostrophe, par MOIGNET (1972a, BFM), BÉDIER (1947, p. 174) et SHORT (2007, p. 152), tandis que SEGRE (1971, p. 398), reproduit dans rolandox, conserve le <e> noté par le ms. (fig. 6.6).

- (6.7) NCA : *il trait almace s'espee de acer brun*
 BFM : *Il trait Almace, s'espee d'acer brun (id. BÉDIER, SHORT)*
 SEGRE : *Il trait Almace, s'espee de acer brun,*
 Dipl. : *I l trait almace fespée de acer brun .*
 « Il tire Almace, son épée d'acier bruni, »
 (*Roland* : rolandox (NCA), roland (BFM), SEGRE 1971, v. 2089, trad. SHORT 1990)

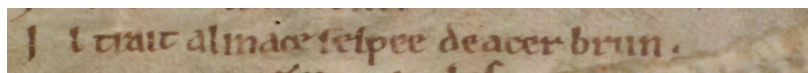


FIGURE 6.6 – *Roland*, ms. d'Oxford, fol. 38r, l. 14 (v. 2089). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

La troisième occurrence, reproduite en (6.8), joue en revanche un rôle métrique pour certains éditeurs : le maintien forcé du <e> de *de* devant initiale vocalique permet de maintenir l'isosyllabisme du vers, qui ferait sans cela 9 syllabes métriques au lieu de 10. Ici, toutes les éditions

4: Dès lors que SEGRE (1971) indique une intervention sur une forme, Dees semble avoir retiré la forme entière de ses données, cf. (6.3) en sec. 6.1.1.

5: Si le scribe du ms. a parfois l'usage de noter le <e> de *de* superscript et lié au le <d>, au point de former un monogramme (SHORT 1973, p. 223-224) — usage nous représentons le cas échéant <d^e>, ce n'est pas le cas de cette occurrence, où <e> est bien placé sur la ligne fondamentale de l'écriture.

6: *Par num d'ocire : au péril de sa vie*; AND « nom », § « par (le) nom de », 3 : *under threat, at the risk of*.

consultées respectent la présence du <e> et, pour expliciter le rôle métrique de la syllabe, SHORT (2007, p. 216) dote ce <e> d'un tréma, convention qui permet de signaler une syllabe maintenue aux seules fins de la métrique.

- (6.8) NCA : *pur co n unt soign de elme ne d osberc*
 BFM : *Pur ço n'unt soign de elme ne d'osberc* (id. BÉDIER, SEGRE)
 SHORT : *Pur ço n'unt soign dë elme ne d'osberc*
 Dipl. : *P ur co nunt foign de elme ne do**l**^{erl}c*.
 « et c'est pourquoi ils n'ont cure ni de heaume ni de haubert »
 (Roland : ro landox (NCA), ro land (BFM), SHORT 1990, v. 3250, trad. SHORT 1990)

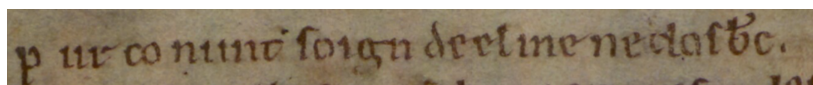


FIGURE 6.7 – *Roland*, ms. d'Oxford, fol. 59r, l. 1 (v. 3250). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

Le cas, ici, est particulier étant donné le statut du mot *elme* (fr.mod. *heaume*), également écrit *helme(s)* avec un <h> initial dans *Roland*. Ce lemme, qui apparaît 42 fois dans *ro landox*, est précédé 12 fois du déterminant *le* ; il n'est jamais immédiatement précédé d'un paroxyton ou d'un autre monosyllabe élidable. Lorsque le mot est sans <h> initial, il déclenche l'élisio de la préposition dans 9 cas (6.9 - 6.17) et ne retient le <e> que dans un seul cas (6.8 *supra*), où le <e> semble également avoir un rôle métrique. Dans les deux occurrences où le mot est doté d'un <h> initial, le <e> est préservé (6.18, 6.19).

- (6.9) *l elme li freint u li carbuncle luisent*
 « [il] lui brise le heaume où brillent des escarboucles »
 (v. 1325, trad. SHORT 2007)
- (6.10) *fiert marganices sur l elme a or agut*
 « [il] frappe [Marganices] sur son heaume doré et pointu »
 (v. 1954, *id.*)⁷
- (6.11) *si l fiert amunt sur l elme a or gemét⁸*
 « il le frappe fort sur le heaume aux gemmes serties dans l'or »
 (v. 1995, *id.*)
- (6.12) *si l fiert en l elme ki gemmēt fut a or*
 « il l'a frappé sur le heaume aux gemmes serties dans l'or ; »
 (v. 2288, *id.*)
- (6.13) *si fiert naimun en l elme principal*
 « et en frappe Naimun sur le sommet du heaume : » (v. 3432, *id.*)
- (6.14) *fier carlemagne sur l elme d acer brun*
 « il frappe Charlemagne sur le heaume d'acier bruni » (v. 3602, *id.*)
- (6.15) *l elme li freint o li gemme reflambent*
 « [il] lui brise le heaume aux gemmes flamboyantes » (v. 3616, *id.*)
- (6.16) *si fiert tierri sur l elme de provence*
 « il frappe Thierry sur le heaume de Provence » (v. 3916, *id.*)
- (6.17) *fiert pinabel sur l elme d acer brun*
 « il frappe Pinabel sur le heaume d'acier bruni » (v. 3926, *id.*)

7: Short édite *l'algalife* au lieu de *Marganices*, et traduit donc « le qualife ».

8: Il faut, suivant MOIGNET (1972a), lire la contraction *sil* de *si* + *il* là où le NCA a *si l* et où Short a *si l'*. *Idem infra*.

- (6.18) *falt li le coer le helme li embrunchet*
 « Le cœur lui manque, son heaume retombe, » (v. 2019, *id.*)
- (6.19) *li amiratz en ad le helme enclin*
 « L'émir en a incliné son heaume » (v. 3504, *id.*)

Le <h> en question correspond au 'h aspiré' de la tradition grammaticale française, c'est-à-dire qu'il est le premier graphème de « mots qui commencent phonétiquement par une voyelle, mais qui d'un certain nombre de points de vue se comportent comme s'ils commençaient par une consonne », dans la mesure où ni l'élision ni la liaison ne sont permises devant ces mots (DELL 1985, p. 262-263). Ce 'h aspiré' est parfois représenté au niveau phonologique comme une consonne (CHAO 1934; BALLY 1944; SCHANE 1967, p. 45-46, 1968, p. 7-8), et plus précisément comme un coup de glotte /ʔ/ (DELL 1973; PAGLIANO 2003) — convention que nous reprenons ici par simple facilité d'usage — ou comme un trait diacritique interdisant la resyllabation en amont de la voyelle initiale (MORIN 1974; CORNULIER 1981; DELL 1985, p. 262). Si l'on suppose parfois que le 'h aspiré' fut véritablement phonétique à une époque ancienne (SCHEER 2015, p. 103) et l'était encore au début de l'ancien français (SÉGÉRAL et SCHEER 2020, GGHF, p. 348-349), les données en (6.9 - 6.17) montrent soit que la situation est plus complexe, soit que ce n'est déjà plus le cas dans *Roland*.⁹ À moins d'admettre que le mot dispose de deux représentations phonologiques, l'une avec et l'autre sans 'h aspiré' (disons, conventionnellement, /ʔəaũmə/ et /əaũmə/), on est forcé d'observer que dans *Roland* le <h> du lemme « heaume » (du lemme, et non de toutes ses formes) n'interdit l'élision que dans une minorité de cas (6.8, 6.18, 6.19) et que, donc, à partir d'une même représentation sous-jacente /ʔəaũmə/, /ʔ/ initial a un effet qui n'est que facultatif. Ici, le scribe semble utiliser le graphème <h> pour indiquer le refus de l'élision, tandis que l'absence de <h> est majoritairement associée à l'élision.

Comparalem et les 'h aspirés'

Pour ne pas polluer nos résultats par la question, orthogonale, des mots subséquents à 'h aspiré' initial, nos requêtes CQP ont été rédigées sans inclure <h> dans la liste des graphèmes consonantiques à l'initiale du mot suivant. En revanche, une telle restriction n'ayant pas été appliquée sur les lemmes, les formes sans 'h aspiré' de lemmes à 'h aspiré' (comme *elme* pour « heaume ») sont bien présentes dans les données, et Comparalem les traite comme des simples contextes subséquents à initiale vocalique. De fait, nos requêtes CQP et, par suite, Comparalem, sont aveugles au lemme du contexte subséquent, et le programme ne peut traiter différemment les mots à initiale vocalique dont le lemme comporte un <h> initial.

Entr. La préposition « entre » apparaît sans <e> dans *Roland* dans 5 occurrences prévocales *entr* (6.20 - 6.24), contre 19 occurrences *entre* devant initiale consonantique (p.ex. 6.25). Il y a donc distribution complémentaire dans ce texte, avec *entr* systématique devant initiale vocalique et jamais devant initiale consonantique ; les éditions concordent sur ce point, le manuscrit étant assez clair pour chaque occurrence, comme illustré en fig. 6.8 pour la première occurrence. Néanmoins, si le défaut de <e> est systématique devant initiale vocalique dans *Roland*, il n'en va pas de même dans l'ensemble du corpus, de nombreux textes ne disposant

DELL (1985). *Les règles et les sons : introduction à la phonologie générative*.

CHAO (1934). « The non-uniqueness of phonemic solutions of phonetic systems ».

BALLY (1944). *Linguistique Générale et Linguistique Française*.

SCHANE (1967). « L'élision et la liaison en français ».

— (1968). « On the abstract character of French « e muet » ».

DELL (1973). *Les règles et les sons : introduction à la phonologie générative*.

PAGLIANO (2003). *L'épenthèse consonantique en français. Ce que la syntaxe, la sémantique et la morphologie peuvent faire à la phonologie*

MORIN (1974). « Règles phonologiques à domaine indéterminé : chute du cheva en français ».

CORNULIER (1981). « H-aspirée et la syllabation des expressions disjonctives ».

SCHEER (2015). *Précis de structure syllabique*.

SÉGÉRAL et al. (2020). « Phonétique Historique ».

9: Il faudrait alors réviser la datation retenue par la GGHF, qui ne note les premières attestations de lemmes à 'h aspiré' sans <h> que du début du XIII^e s. (SÉGÉRAL et SCHEER 2020, GGHF, p. 348-349).

que de formes sans défaut de <e>, ou disposant de formes avec et sans défauts devant initiale vocalique, comme indiqué en fig. 6.3 *supra*.

- (6.20) *entr els en unt e orgoil e cunfort*
 « ils se rassurent et leur orgueil croît »
 Dipl. : *entrelf en unt 7 orgoil 7 cunfort*.¹⁰
 (Roland, v. 1941, trad. SHORT 2007)

10: Nous rendons par '7' la forme archaïque de l'esperluette, abréviation de *et*.

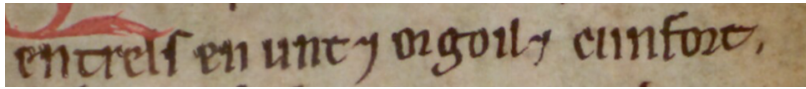


FIGURE 6.8 – *Roland*, ms. d'Oxford, fol. 35v, l. 1 (v. 1941). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

- (6.21) *asez est fols ki entr els se [dementet]*
 « Entouré d'eux, seul un fou s'inquiéterait. »
 (Roland, v. 3010, trad. *id.*)
- (6.22) *entr els nen at ne pui ne val ne tertre*
 « Ni mont, ni val, ni tertre ne les séparent »
 (Roland, v. 3292, trad. *id.*)
- (6.23) *bataille veit cil ki entr els volt estre*
 « Celui qui lutte parmi eux voit ce qu'est une vraie bataille. »
 (Roland, v. 3404, trad. *id.*)
- (6.24) *enmi le camp amdui s entr encuntrentent*
 « tous deux se rencontrèrent au milieu du champ »
 (Roland, v. 3567, trad. *id.*)
- (6.25) *entre les helz ad plus de mil manguns*
 « la garde [de l'épée sus-mentionnée] en vaut plus de mille mangons »
 (Roland, v. 620, trad. *id.*)

Sor. Quant à « sor », le lemme n'apparaît dans *Roland* qu'en défaut de <e> : une seule fois devant initiale vocalique (6.26) et 67 fois devant initiale consonantique (p.ex. 6.27), et l'occurrence pré-vocalique se trouve dans un vers syntaxiquement et sémantiquement correct mais hypométrique. Il faut se garder de conclure à une quelconque restriction due au contexte pré-vocalique : on trouve également 10 occurrences *sur un*, telles que celle reproduite en (6.28) ; ces occurrences ont échappé à nos résultats à cause du caractère trop exigeant de l'exclusion des <i>, <u> et <v> initiaux dans nos requêtes CQP. Par ailleurs, si la seule occurrence pré-vocalique repérée par Comparalem se trouve dans un vers qui demande une intervention éditoriale, celle-ci ne semble pas devoir changer le contexte subséquent de *sur* : si l'on retient la correction proposée par Segre et reprise par Short, c'est après *altre* qu'il faut intervenir pour restituer ce qui manque dans le ms. (reproduit en fig. 9.12).

- (6.26) NCA : *un mort sur altre geter* (id. BÉDIER, BFM)
 SEGRE : *Un mort sur altre <a la tere> gete[r]*
 SHORT : *Un mort sur altre a la tere geter*
 Dipl. *vn mort fur altre geter.*
 « jeter les mots les uns sur les autres »
 (Roland : rolandox (NCA), SEGRE 1971, SHORT 1990, v. 1970, trad. SHORT 1990)

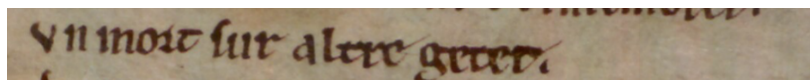


FIGURE 6.9 – *Roland*, ms. d'Oxford, fol. 36r, l. 8 (v. 1970). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

- (6.27) *e cez escuz sur cez helmes cruisir*
 « et les écus grincer sur les heaumes »
 (rolandox, v. 3485, trad. SHORT 1990)
- (6.28) *sur un perrun de marbre bloi se culchet*
 « là il s'allonge sur un bloc de marbre gris »
 (rolandox, v. 13, trad. SHORT 1990)

6.2.3.2 Lemmes attestés en défaut de <e> au cours du XII^e s.

Nos trois autres lemmes apparaissent en défaut de <e> à partir de la moitié du XII^e s. « Encontre » apparaît sous la forme *encontr* au milieu du siècle, suivi de « contre » (sous la forme *contre*) dix ans plus tard, et de *desor* cinq ans plus tard.

Encontr. Le troisième lemme à apparaître en défaut de <e> est « *encontr* », dans deux versions du *Charroi de Nîmes*, au milieu du XII^e s. : *nimb1* et *nimb2*, avec une attestation en défaut de <e> pour chaque texte. Sur le graphique en (6.3), les deux points pour « *encontr* » de *nimb1* et *nimb2* sont superposés à 100%, dans la mesure où aucun de ces textes ne contient d'autre attestation du lemme « *encontr* ». En revanche, chaque texte contient deux attestations du lemme « *contre* », dont le <e> est bien noté.¹¹ Les deux attestations *encontr* sont placées devant initiale vocalique, telle que produit en (6.29 - 6.30), tandis que les quatre occurrences de *contre* sont devant initiale consonantique, aussi il n'est pas possible de dissocier ce qui tient de la spécificité du lemme préfixé ou du contexte pour ces premières attestations. Les autres versions du *Charroi de Nîmes* ne contiennent pas de passage syntaxiquement équivalent qui pourrait donner lieu à une comparaison.

- (6.29) *voit les le roi incontr eus s est levez*
 « le roi les voit et s'est levé à leur *encontr* »
 (*nimb1*, p. 13, trad. pers.)
- (6.30) *voit les le roy incontr eus s est levez*
 « le roi les voit et s'est levé à leur *encontr* »
 (*nimb2*, p. 11, id.)

nimb1 : ANON. (1150*). *Le charroi de Nîmes*. Ms. B1; Dees : Aisne; compo. nil, ca 1150; ms. Paris, ca 1335; MS1 (transcr. inédite L. Schoesler).

nimb2 : ANON. (1150*). *Le charroi de Nîmes*. Ms. B2; Dees : Normandie; compo. nil, ca 1150; ms. Paris, ca 1335; MS1 (transcr. inédite L. Schoesler).

11: *nimb1* : p. 11, 12; *nimb2* : p. 10.

eneas : ANON. (1160*). *Éneas*. Ms. A; Dees : Haute-Marne; compo. norm., ca 1160; ms. lorr., ca 1200; CR2 (SALVERDA DE GRAVE 1964-68).

Contr. Le défaut de <e> sur « contre » apparaît 10 ans plus tard, dans *eneas*. Ce texte ne contient que deux attestations du lemme : l'une en défaut de <e> devant initiale vocalique (6.31), l'autre sans défaut de <e> devant initiale consonantique (6.32). Le texte ne contient aucune occurrence du lemme « encontre ».

- (6.31) *eneas de bien loin les vit contr els ala*
« Énée les vit au loin ; il alla à leur rencontre »
(*eneas*, p. 13, trad. pers.)
- (6.32) *contre lui est dido venue*
« Didon est venue à sa rencontre »
(*id.*, p. 14, *id.*)

Desor. Cinq ans plus tard, « desor » apparaît en défaut de <e>, dans *lanvals* (6.33), avant d'être également attesté dans *pap*, *fablesC*, *fablesK*, *perpraag*, *romd* et *fab4e* (6.34-6.38), textes dont la composition est datée entre 1175 et 1200.

- (6.33) *deus puceles voient uenir desors parlefroiz*
« Ils voient venir deux jeunes filles sur un palefroiz »
(*lanvals*, p. 9, trad. pers.)
- (6.34) *desur nos est aposé*
« il nous est imposé »
(*pap*, p. 47, *id.*)
- (6.35) *s acist desor le suel*
« il s'assit sur le seuil »¹²
(*fablesC*, p. 3, *id.*)
- (6.36) *vns gourpis vint dessour sel vit en haut parla*
« un renard vint au-dessus ; il [le pigeon] lui parla à haute voix/en hauteur »
(*fablesK*, p. 34, *id.*)
- (6.37) *desor le lit ot estendue une grant couste*
« sur le lit était étendue une grande couette »
(*perpraag*, p. 1, *id.*)
- (6.38) *tant y musai que sor mon coute m endormi tres desor la coute insi icelui ior m auint*
« je m'attardais tant sur ma couette que je m'endormis profondément sur ma couette ; c'est ainsi que je me réveillais ce jour-là »
(*fab4e*, p. 13, *id.*)

Ce lemme n'apparaît, annoté comme une préposition et dans un contexte auquel Comparalem est sensible, que dans huit occurrences, dont une sans défaut de <e>.¹³ Six occurrences sur sept de « desor » en défaut de <e> sont donc concentrées entre 1175 et 1200. La dernière occurrence de « desor » prépositionnel en défaut de <e> apparaît dans *pseuturp*, un siècle plus tard (6.39). L'unique occurrence du lemme « desor » sans défaut de <e> apparaît dans *fablesK* (6.40) ; ce texte possède donc les deux variantes.

lanvals : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. S ; Dees : nil ; compo. N.-O., ca 1165 ; ms. frc., ca 1300 ; MS2 (MARIE DE FRANCE 1958) ; DEAF : MarieLanvR².

pap : ANON. (1175*). *La vie du pape Saint Gregoire*. Ms. A1 ; Dees : nil ; compo. Nord, ca 1175 ; ms. tour., ca 1238 ; MS1 (SOL 1977) DEAF : SGreg^a1S.

12: Ou le « sur le sol », le contexte (il entre dans un moulin) étant ambigu.

fablesC : MARIE DE FRANCE (1180*). *Fables*. Ms. C ; Dees : nil ; compo. NO, ca 1180 ; ms. lorr., ca 1275 ; MS (transcription inédite).

fablesK : MARIE DE FRANCE (1180*). *Fables*. Ms. K ; Dees : nil ; compo. NO, ca 1180 ; ms. nil, ca 1310 ; MS (transcription inédite).

perpraag : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS Prague ; Dees : nil ; compo. Troyes, ca 1180 ; ms. Nord Est, ca 1300 ; MS1 (transcr. inédite R. L. H. Lops).

fab4e : ANON. (1200*). *Fabliau n°4*. Ms. E ; Dees : Nièvre, Allier ; compo. pic., ca 1200 ; ms. frc., ca 1290 ; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).

13: Une occurrence annotée prépositionnelle, et en défaut de <e>, apparaît dans un contexte exclu de nos requêtes : devant le chiffre latin .i. dans *fablesT*, p. 30-31 : *porta le sanc en sa chambre deseur .i. banc*, « il répandit le sang dans sa chambre, sur un banc ».

fablesT : MARIE DE FRANCE (1180*). *Fables*. Ms. T ; Dees [Angleterre] ; compo. NO, ca 1180 ; ms. champ., ca 1287 ; MS (transcription inédite).

- (6.39) *et vis que un jor qu il traspasset de leu en autre, que li jors devient toz obscurs et noirs en desoures, et grant flaimde de feu traspassai a senestre per devant ses eouz*
 « et je vis [Charlemagne] traverser de part en part [le pont sur le Rhin], à ce moment le jour se fit tout à fait obscur et noir sous [le pont], et une grande flamme paraissait à gauche devant ses yeux »¹⁴
 (pseuturp, p. 55, trad. pers.)
- (6.40) *dist dou lou et de l aingnel qui beuoient a zz ruissel li leus au desseure beuoit et li aingniax aual estoit*
 « [l']histoire du loup et de l'agneau qui buvaient dans un ruisseau, le loup buvait en amont [au-dessus] et l'agneau était en aval »
 (fablesK, p. 2, id.)

14: Le passage raconte la vision annonçant la mort de Charlemagne. Les crochets restituent le contenu des anaphores selon le contexte qui précède.

pseuturp : ANON. (1290*). *The burgundian translation of the Pseudo-Turpin Chronicle*. Dees : Haute-Saône; compo. bourg., ca 1290; ms. *idem*; MS (WALPOLE 1949, 1948); DEAF : Turpin⁷W.

— **Un lemme, plusieurs attributs de POS.** Les formes du lemme « desor » sans défaut de <e> ne se réduisent pas à cette seule occurrence : 187 formes avec <e> sont annotées comme étant des adverbes (pos="ADV" ; p.ex. 6.41 a). Elles demeurent néanmoins minoritaires, puisque 438 formes en défaut de <e> sont également annotées comme des adverbes (p.ex. 6.41 b et 6.41 c).

- (6.41) « Desor » non prépositionnel (pos != "PRE")
- (a) *li lis fu fais d un estrain blanc ou il ot grant keute de plume et pour ce que on ne s i emplume ot desseure vne keute pointe aubree*
 « le lit était fait d'une litière blanche sur laquelle il y avait une grande couette de duvet et, pour qu'on ne se retrouve pas couverts de plumes, il y avait par-dessus une [housse de ?] couette finement ouvragée »
 (fab4f, p. 5, trad. pers.)
- (b) *[.i.] corbel s estoit assis desor*¹⁵ *lou dos d une berbis*¹⁶
 « un corbeau s'était assis sur le dos d'une brebis »
 (fablesC, p. 41, id.)
- (c) *vns corbiar estoit assis desseur*¹⁷ *le dos d une berbis*
 « un corbeau était assis sur le dos d'une brebis »
 (fablesK, p. 32, id.)

fab4f : ANON. (1200*). *Fabliau n°4*. Ms. F; Dees : Somme, Pas-de-Calais; compo. pic., ca 1200; ms. nil, ca 1300; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).

15: pos="ADV" & taggerpos="ADV".

16: Le nombre [.i.] est manquant dans le NCA, remplacé par zz.

17: pos="ADV" & taggerpos="ADV".

Variables de POS du NCA3 et opérateur relationnel

Nous faisons ici référence aux variables du NCA3 et utilisons l'opérateur relationnel '!=':

pos correspond à l'étiquette de partie du discours de Dees explicitée lors de la refonte du corpus;

taggerpos correspond à l'étiquette proposée par Tree Tagger;

!= correspond à l'opérateur relationnel *est différent de*, équivalent de l'opérateur mathématique '≠'.

Ainsi, le lemme « desor » peut être annoté comme une préposition ou comme un adverbe, et l'est majoritairement en tant qu'adverbe. Dans la mesure où l'annotation de la partie du discours, ici dépendante d'une analyse contextuelle, humaine (pos) ou mécanique (taggerpos), peut indiquer une information pertinente pour la phonologie en termes

de domaines, il ne s'agit pas de l'ignorer absolument. Néanmoins, la comparaison des exemples où « desor » est annoté PRE et de ceux où il est annoté ADV montre une certaine fluctuation. Ainsi, en (6.35) comme en (6.41 b, c), « desor » est immédiatement précédé du même verbe (*s'asseoir dessus qqch.*), et si le plus-que-parfait (*estoit assis*) en (6.41 b, c) peut éventuellement induire Tree Tagger à considérer que c'est un participe et non un verbe conjugué qui précède « desor », l'analyse syntaxique de « desor » demeure la même dans les trois exemples. L'annotation en partie du discours de ce lemme semble donc mal assurée. Si nous n'intégrons pas, ici, les occurrences notées comme adverbiales, il convient néanmoins de conserver une posture précautionneuse face à ces données.

Il n'en demeure pas moins que le lemme « desor » est sujet à un défaut de <e> majoritaire, attesté dès la seconde moitié du XII^e s. dans les formes annotées comme des prépositions. Nous n'avons pas trouvé une erreur d'annotation d'une telle ampleur pour « sor »; d'une part les occurrences de « sor » non prépositionnelles sont peu nombreuses (154 occ. pos!="PRE" vs 3490 occ. pos="PRE"), et d'autre part une bonne partie d'entre elles concerne soit des occurrences ambiguës soit des occurrences qui ne tiennent clairement pas du lemme prépositionnel, comme lorsque le mot lemmatisé « sor » représente le lemme nominal « suer » (6.42 a) ou le lemme adjectival désignant la robe d'un cheval brun-roux (6.42 b). La qualité des annotations fondées sur les étiquettes de Dees (pos) permet, dans ces exemples, de distinguer la préposition du nom et de l'adjectif, alors que l'annotation en parties du discours produite par Tree Tagger s'en tient à supposer que ces occurrences sont des prépositions, comme indiqué en notes sur ces exemples.

(6.42) « sor » != "PRE" peut représenter un autre lemme

- (a) *bele sor*¹⁸ *coment e pur quei jo me mor*
« Chère sœur, comment et pourquoi suis-je en train de mourir ? »
(*ipo*, p. 50, trad. pers.)
- (b) *et sist sour un palefroi sor*,¹⁹
« et [il] se tenait sur un palefroi brun-roux, »
(*amad*, p. 39, *id.*)

18: pos="NOM_suj_femi_sg" & taggerpos="PRE".

ipo : HUE DE ROTELANDE (ca 1185). *Ipomedon*. Compo. a.n., ca 1185; ms. a.n., ca 1190; CR (transcription inédite).

19: *sor* pos="ADJ_obj_masc_sg" & taggerpos="PRE".

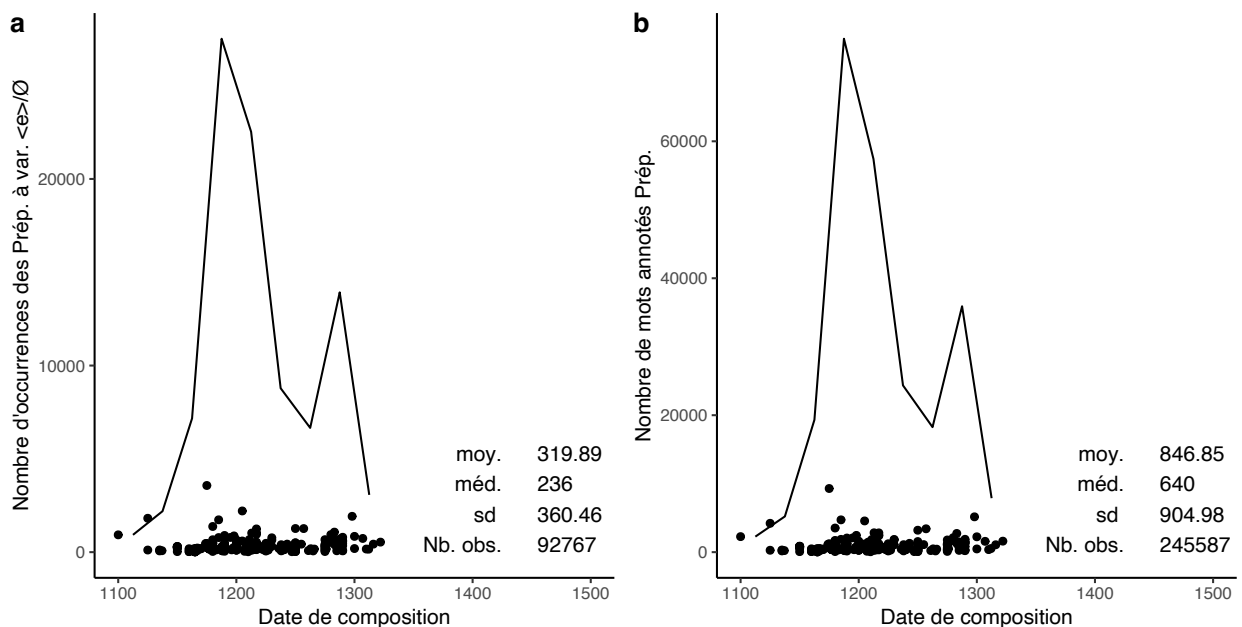
amad : ANON. (1210*). *Amadas et Ydoine*. Ms. P; Dees : Aisne; compo. a.n., ca 1210; ms. pic., ca 1289; CR3 (ANON. 1974); DEAF : AmYdR.

Par ailleurs, si les occurrences de « sor » non prépositionnelles ne semblent pas poser problème, ne serait-ce que par leur quantité marginale, les occurrences de « sor » prépositionnelles semblent bien correspondre à la préposition « sor ». Étant donné le nombre très élevé d'occurrences (3490 occ.), nous n'avons pas été en mesure de les vérifier manuellement, en contexte, une par une, mais nous avons vérifié les 100 occurrences sans défaut de <e>. Si ces 100 occurrences sont bien prépositionnelles et représentent bien le lemme « sor », on y trouve un certain nombre d'usages absolus de la préposition, notamment dans l'expression *corir sur* (AND : « to run over, onto ») en (6.43). Sur les 3484 occurrences en défaut de <e>, nous avons effectué des sondages, qui n'ont pas révélé de problème d'annotation, mais la sensibilité de ces sondages n'est pas assurée, étant limitée à quelques centaines de formes.

- (6.43) *lor courent seure mout en ocient*
« ils passent à l'assaut, ils en tuent beaucoup »
(*athi*, p. 42, trad. pers.)

athi : ANON. (1200*). *Li romanz d'Athis et Prophilius*. Dees : Aisne; compo. pic., ca 1200; ms. frc., ca 1350; CR2 (HILKA 1912-1916); DEAF : AthisH.

En somme. Tous nos lemmes sont atteints par le défaut de <e> dès le premier siècle du corpus et, pour « de », « entre » et « sor », dès le premier texte, *Roland*, daté de ca. 1100. Il peut paraître notable que le lemme « entre » est attesté sous sa forme *entr* dès *Roland* tandis que les attestations *contr* et *encontr* n'apparaîtront qu'à partir du milieu du XII^e s., mais « entre » (1854 occ.) est presque deux fois plus courant que « contre » (1005 occ.) et « encontre » (26 occ.), et le corpus ne dispose que de cinq textes dans la première moitié du XII^e s. Comme le défaut de <e> du lemme « contre » demeure minoritaire devant initiale vocalique, et que le lemme « encontre » est tout simplement trop rare, ce décalage entre d'une part « entre » et d'autre part « contre » et « encontre » ne paraît donc pas significatif. Il ne semble donc pas y avoir de phénomène de diffusion de la chute de la voyelle finale au sein du corpus lexical des prépositions ; non seulement les élisions sont déjà présentes dans *Roland*, mais l'élision y est parfois plus systématique (cf. distribution complémentaire de *entr*). De fait, nos lemmes sont attestés en défaut de <e> soit dès *Roland* soit dans la seconde moitié du XII^e s., qui correspond au pic dans la taille du corpus, telle que représentée en fig. 6.10. Sur cette figure, la courbe en (a) cumule le nombre d'occurrences des cinq lemmes sélectionnés, sur une fenêtre de 25 ans, et est strictement parallèle à la courbe en (b), qui cumule les occurrences de toutes les prépositions. Nos formes défectueuses sont donc attestées avant le premier pic dans la taille du corpus, que celle-ci soit envisagée uniquement pour nos cinq lemmes (a) ou pour l'ensemble des prépositions (b). Les défauts de <e> des prépositions ne semblent donc pas attendre que le corpus grossisse pour s'exprimer.



Chaque point représente un texte.

Les courbes relient la somme des occurrences sur une fenêtre de 25 ans.

FIGURE 6.10 – Distribution diachronique des occurrences des lemmes de prépositions à défaut de <e> (a), en regard de la distribution diachronique des occurrences de tous les lemmes de prépositions.

6.2.3.3 Fréquence des lemmes et défauts de <e>, ou plutôt typologie des défauts

Du fait du faible nombre de lemmes retenus, l'analyse de l'effet de la fréquence lexicale ne peut être conclusive. Le graphique en fig. 6.11 représente le taux de défaut de <e> de chaque lemme en fonction de sa fréquence (occurrences avec et sans défaut de <e>), et les données sur lesquelles il se fonde sont exposées en table 6.2, ce qui permet d'identifier chaque point.

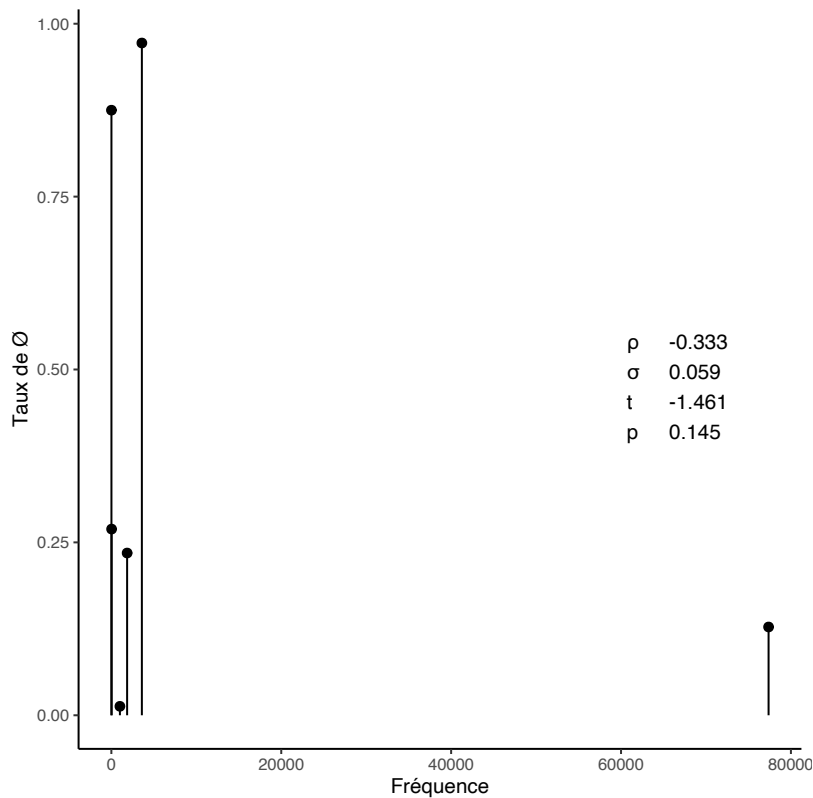


FIGURE 6.11 – Distribution des lemmes en fonction de leur fréquence et de leurs taux de défaut de <e>, prépositions.

TABLE 6.2 – Relation entre fréquence et taux de défaut de <e>, pour chaque lemme de préposition.

lemma	nb.Ø	nb.E	frq.	tx.Ø
de	9878	67 491	77 369	13%
sor_relemma ^a	3490	100	3590	97%
entre_relemma	435	1419	1854	23%
contre	13	992	1005	1%
encontre_relemma	7	19	26	27%
desor ^a	7	1	8	88%

a. Ne contient pas les occ. dont pos!="PRE".

Si l'on observe bien, comme pour les adjectifs féminins singuliers, une concentration des taux les plus élevés autour des basses fréquences (ici < 3600, à comparer aux plus de 77 000 occurrences du lemme « de »), cet effet est probablement un artéfact, dû au fait qu'aucun lemme prépositionnel n'a une fréquence contenue entre celle de « sor » (3590 occ.) et celle du lemme « de » (77 369 occ.). Au sein même des lemmes à 'basse fréquence', un lemme sur cinq, « contre », ne s'inscrit d'ailleurs pas dans cette tendance (il correspond au point proche de l'intersection des axes en fig. 6.11). En fonction de ce qui a été détaillé *supra* lors de la description de l'ordre d'apparition en défaut de <e> des lemmes, il semble plutôt que le seul facteur pertinent ici soit la sensibilité au contexte subséquent, dans la mesure où « de », « entre », « contre » et « rencontre » n'ont de défauts de <e> que devant initiale vocalique (avec quelques rares exceptions pour « de »). La fréquence d'emploi des lemmes ne semble donc pas avoir d'effet ici.

6.2.3.4 Fréquence, collocations et défauts de <e>

Étant donnée la systématisme presque parfaite des défauts de <e> de « de » devant initiale vocalique (cf. sec. 6.2.4.1 *infra*), il n'y a pas lieu de rechercher un quelconque effet de fréquence de *token* pour ce lemme, ni d'effet de fréquence de *type* étant donné que le déterminant est toujours suivi du groupe nominal qu'il détermine. Nous n'avons pas non plus trouvé de construction spécifique pour « sor » et « desor ». En revanche, il existe des constructions spécifiques qui, avec « entre », « contre » et « rencontre », sont régulièrement associées à des défauts de <e>.

Entr. Sur les 435 occurrences de la préposition « entre » en défaut de <e>, 393²⁰ se trouvent devant le pronom personnel masculin pluriel au cas régime « eux » (p.ex. 6.44), 17 occurrences²¹ se trouvent devant le même pronom mais au féminin (p.ex. 6.45), et 23 occurrences se trouvent devant un verbe, que celui-ci soit conjugué (p.ex. 6.46), à l'infinitif (p.ex. 6.47) ou au participe passé passif (p.ex. 6.48).

- (6.44) *lors parolent entr eus et dient*
« alors ils discutent entre eux, puis disent [...] »
(lin, p. 3, *id.*)
- (6.45) *entr eles s estoit mis*
« il s'était placé entre elles »
(chret1, p. 9, *id.*)
- (6.46) *enfanz qui s entr amerent*
« des jeunes gens qui s'aimaient l'un l'autre »
(deusamS, p. 1, trad. pers.)
- (6.47) *sanz grant mal n i pooient ansamble antrer n anmi la porte entr ancontrer*
« ils ne pouvaient sans grande difficulté entrer ensemble sans se heurter l'un l'autre au milieu de la porte »
(yvh, p. 19, *id.*)
- (6.48) *se sont entr acolé tant fort se baisent et estraingnent*
« ils se sont enserré si fort, ils s'embrassent et s'étraignent »
(amile, p. 5, *id.*)

20: 392 occ. où « eux » est bien annoté `pos="PRO_pers_acc_3_masc_pl"` et une occurrence *entre aus* où *aus* est annoté comme la contraction d'une préposition et d'un déterminant (`pos="PREDET_a_obj_masc_pl"`; *hista*, p. 42), où il s'agit visiblement d'une erreur d'annotation.

21: 16 occ. où « eles » est bien annoté `pos="PRO_pers_acc_3_femi_pl"` et une occurrence où *eles* est annoté comme un pronom personnel sujet (`pos="PRO_pers_acc_3_femi_pl"`), probablement à cause de l'omission du sujet réel : *etr eles dirent e consillierent*, « elles dirent et décidèrent entre elles » (*fablesK*, 14.)

lin : ROBERT DE BORON (1213*). *Merlin*. Ms. A; Dees : Yonne; compo. Sud-Est, ca 1213; ms. bourg., ca 1250; CR2 (MICHAM 1980); DEAF : MerlinM.

chret1 : CHRÉTIEN DE TROYES (1180*). *Le Chevalier au lion & Perceval*. Dees : Hainaut; compo. champ. mérid., ca 1180; ms. champ., ca 1238; MS (transcr. inédite R. L. H. Lops); contenu présent ailleurs dans le corpus : incipit du *Chevalier au lion* (v. 1-998) dans yvf, et *Perceval* (v. 69-2000) dans perr.

deusamS : MARIE DE FRANCE (1165*). *Les Deus Amanz*. Ms. S; Dees : nil; compo. a.n., ca 1165; ms. frc., ca 1300; MS (transcr. inédite sans transcripteur indiqué).

yvh : CHRÉTIEN DE TROYES (1177*). *Le Chevalier au lion*. Ms. H; Dees : nil; compo. champ. mérid., ca 1177; ms. champ., ca 1235; MS (transcr. inédite R. L. H. Lops); également contenu dans chret2.

amile : ANON. (1200*). *Ami et Amile*. Dees : Haute-Marne; compo. nil, ca 1200; ms. lorr., ca 1275; MS1 (DEMBOWSKI 1969); DEAF : AmAmD.

Seules deux occurrences *entr* se trouvent devant un autre mot, en l'occurrence le pronom (annoté comme un adjectif) *autres* en (6.49) et le nom « or » en (6.50).

- (6.49) *il ne se deigna [...] n a sei sozmetre ne ses deus mains entr autres metre*
« il ne deigna ni à se soumettre ni à placer ses deux mains dans celles de quelqu'un d'autre »
(benoit, p. 46, *id.*)
- (6.50) *vint somiers chargiés entr or fin et argent*
« vingt bêtes de somme chargées d'or fin et d'argent »
(flo, p. 6, *id.*)

Parallèlement, la construction *entre* + PROPER3pl CR (masc. ou fém.) n'est attestée sans défaut de <e> que dans 137 occurrences, tandis qu'on ne trouve aucune occurrence, avec ou sans défaut de <e>, de *entre ele* au singulier, la sémantique de la préposition ne le permettant pas. Enfin, il faut noter que le pronom personnel « eux/elles » est le seul pronom personnel pluriel dont l'initiale soit vocalique, ce qui réduit fortement l'éventail de contextes « entre » + PROPER dans lequel « entre » est susceptible de connaître un défaut de <e>.

Cet effet très marqué doit être nuancé par les possibilités combinatoires autorisées par la préposition *entre*; celle-ci ne se trouve associée à un élément non pronominal à initiale vocalique que dans 77 occurrences (dont 6.49 et 6.50 *supra*). Néanmoins, il est notable que les 395 occurrences du type *entr eus* ne s'opposent qu'à 125 occurrences sans défaut de <e> (type *entre eus*), et que l'on retrouve le même effet, quoique moins marqué, au féminin, avec 8 occurrences du type *entre eles* pour 17 occurrences du type *entr eles*. Aussi, si le faible nombre d'attestations de la préposition « entre » avec d'autres éléments que le pronom personnel objet pluriel au cas régime ne permet pas de définir précisément quelle est la sélectivité de l'association, il demeure notable que la construction « entre » + PROPER3pl CR est majoritairement en défaut de <e>, avec un effet plus marqué au masculin qu'au féminin. « Entre » semble donc privilégier son inclusion au sein du domaine du mot suivant lorsque celui-ci est une pronom personnel régime; autrement dit, le syntagme pronominal prépositionnel semble induire l'inclusion de la préposition dans le domaine du pronom.

Contr. Sur les 13 occurrences *contr*, six se trouvent également devant le pronom personnel objet « eux » (6.51 - 6.56). Étant donné que ces occurrences représentent près de la moitié des occurrences défectueuses *contr*, il y a lieu de supposer que l'association entre « contre » et « eux » est de nature à favoriser le défaut de <e> du premier terme.

- (6.51) *qui arme osoct contr eus saisis*
« qui [osa ?] prendre arme contre eux »
(benoit, p. 17, trad. pers.)
- (6.52) *nul plus felon ne plus engres n en out contr eus set devers nos*
« il n'y avait personne de plus traitée ni de plus violent contre eux ou envers nous »
(chro, p. 25, *id.*)

benoit : BENOIT DE SAINTE-MAURE (1174*). *Chronique des ducs de Normandie*. Ms. T; Dees : Vendée, Deux-Sèvres; compo. poit., ca 1174; ms. poit., ca 1190; MS1 (BENOIT DE SAINTE-MAURE 1951-1967).

flo : ANON. (1213*). *Florence de Rome*. Dees : Nièvre, Allier; compo. pic., ca 1213; ms. Est, ca 1300; CR1 (WALLENSKÖLD 1907-1909); DEAF : FlorenceW.

benoit : BENOIT DE SAINTE-MAURE (1174*). *Chronique des ducs de Normandie*. Ms. T; Dees : Vendée, Deux-Sèvres; compo. poit., ca 1174; ms. poit., ca 1190; MS1 (BENOIT DE SAINTE-MAURE 1951-1967).

chro : BENOIT DE SAINTE-MAURE (1174*). *Chronique des ducs de Normandie*. Ms. [Tours, Bibl. Munic. 903]; Dees : [Vendée, Deux-Sèvres]; compo. poit., ca 1174; ms. a.n., ca 1217; CR2 (BENOIT DE SAINTE-MAURE 1951-1967); DEAF : BenDucF.

- (6.53) *charles li chalf qui rei fu de france est tost contr els uenu*
« Charles le Chauve qui fut roi de France est aussitôt allé à leur rencontre »
(edmund, p. 65, *id.*)
- (6.54) *eneas de bien loin les vit contr els ala*
« Énée les vit au loin ; il alla à leur rencontre »
(eneas, p. 13, trad. pers.)
- (6.55) *lors ont un chevalier veu qui lou chemin venoit contr eus li chevaliers venoit toz seus*
« alors ils ont vu un chevalier qui venait dans leur direction, le chevalier s'avançait tout seul »
(epee, p. 18-19, *id.*)
- (6.56) *lor anemi ne se porent contr eux tenir*
« leurs ennemis ne purent tenir contre eux »
(mace, p. 71, *id.*)

En revanche, aucune autre occurrence de « contre » en défaut de <e> n'apparaît devant un autre pronom personnel, mais dans notre corpus les associations de « contre » avec un pronom à initiale vocalique sont rares : *contre eles* apparaît trois fois (guigS, p. 15; hista, p. 43, lanvalP), *contre autrui* apparaît deux fois (beauv, p. 68; contre, p. 42), et *contre yceuls* n'apparaît qu'une fois (cont re, p. 42). En comparaison, « contre eux » apparaît 26 fois sans défaut de <e>, comme illustré en (6.57 - 6.59); le syntagme est donc attesté 32 fois dans le corpus, et le défaut de <e> n'y est que minoritaire, sans être marginal pour autant. En revanche, le défaut est presque marginal à l'échelle de l'ensemble des attestations du lemme : « contre » est attesté 1005 fois.

- (6.57) *quant les dames les virent venir elles sont levees contre eulz*
« quand les dames les virent venir, elle se sont levées à leur rencontre »
(cass, p. 15, trad. pers.)
- (6.58) *elle vint contre eulz*
« elle vint à leur rencontre »
(helc, p. 22, *id.*)
- (6.59) *vous vainterés les crestiens a l heure que contre aus irés*
« vous vaincrez les chrétiens lorsque vous irez les affronter »
(nic, p. 5, *id.*)

Trois autres occurrences défectueuses *contr* apparaissent devant les mots « aval » et « amont » (6.60 - 6.62), dans une construction proche des constructions *contre munt* (« upstream, upwards ») et *contre val* (« downstream, downwards, down ») attestées par l'AND.

- (6.60) *la pucelle desvaule contr aval le plainchie*
« la jeune fille dévalle sur le plancher »
(floov, p. 15, trad. pers.)
- (6.61) *lors s estoit regardez tot contr avaul zz chan et vit venir*
« alors il regardait un champ en contre bas et il vit venir [qqun] »
(floov, p. 26, *id.*)
- (6.62) *si comensa a monter contr amont une eschiele*
« et il commença à grimper sur une échelle »
(hista, p. 104, *id.*)

edmund : DENIS PIRAMUS (1195*). *La Vie Saint Edmund le Roi*. Dees : [Angleterre]; compo. a.n., ca 1195; ms. a.n., ca 1325; MS2 (DENIS PIRAMUS 1935); DEAF : EdmK.

eneas : ANON. (1160*). *Éneas*. Ms. A; Dees : Haute-Marne; compo. norm., ca 1160; ms. lorr., ca 1200; CR2 (SALVERDA DE GRAVE 1964-68).

epee : ANON. (1200*). *Le chevalier à l'épée*. Dees : Nièvre, Allier; compo. bourg., ca 1200; ms. bourg. sept., ca 1310; MS1 (JOHNSTON et OWEN 1972); DEAF : ChevEspJ.

mace : MACÉ DE LA CHARITÉ (1300*). *La Bible*. Dees : Nièvre, Allier; compo. Centre, ca 1300; ms. Berry, 1343; CR (MACÉ DE LA CHARITÉ 1977); DEAF : BibleMacéV.

cass : ANON. (1275*). *Le Romand de Cassidorus*. Dees : Marne; compo. pic., ca 1275; ms. frc., ca 1290; CR2 (PALERMO 1963); DEAF : CassidP.

chro : BENOIT DE SAINTE-MAURE (1174*). *Chronique des ducs de Normandie*. Ms. [Tours, Bibl. Munic. 903]; Dees : [Vendée, Deux-Sèvres]; compo. poit., ca 1174; ms. a.n., ca 1217; CR2 (BENOIT DE SAINTE-MAURE 1951-1967); DEAF : BenDucF.

nic : JEHAN BODEL (1197*). *Le jeu de Saint Nicolas*. Dees : Somme, Pas-de-Calais; compo. art., ca 1197; ms. Arras, ca 1295; MS1 (JEHAN BODEL 1965); DEAF : BodeINicJ².

floov : ANON. (1190*). *Floovant, Chanson de Geste du XIII^e siècle*. Dees : Bourgogne; compo. Sud-Est, ca 1190; ms. bourg., ca 1325; MS1 (ANDOLF 1965); DEAF : FloovA.

hista : WAUCHIER DE DENAIN (1213*). *Histoire ancienne jusqu'à César (Estoires Rogier)*. Ms. [BFN fr. 20125 & Vienne Nat. Bibl. 2576]; Dees : [Haute-Marne]; compo. Belgique, ca 1213; ms. Nord-Est, ca 1287; MS1 (WAUCHIER DE DENAIN 1995); DEAF : HistAncV.

Si trois attestations, dont deux proviennent du même texte, sont insuffisantes pour déduire un effet de la fréquence d'association (*type* voire *token*) de ces mots, il faut noter que les formes non défectueuses ne sont également que très faiblement attestées dans ces constructions : nous n'avons trouvé aucune construction *contre aval* sans défaut de <e>, et une seule construction *contre amont* sans défaut de <e> (6.63).

- (6.63) *si aveit une pierre de ancianeovre de sarrazins faite e polie [...] si est droite contre amont*
 « il y avait une pierre, ancienne construction polie des Sarrasins [...] et elle était toute verticale [droite en hauteur] »
 (turp, p. 8, *id.*)

Par ailleurs, l'AND et le GD attestent de formes en un seul mot, sous l'entrée « contre » (AND : *cuntreval*), ou plus souvent sous une ou plusieurs entrées distinctes (GD : « contraval », « contreval », « contremont » ; AND : « contremont »). De fait, l'élément *contre* est très productif en tant que préfixe, également dans la production de verbes,²² de noms,²³ d'adjectifs.²⁴ Il semble donc que, lorsque les conditions syntaxiques et fréquentielles sont réunies, la frontière entre *contre* et le mot suivant puisse s'effacer ce qui, lorsque le mot suivant commence par une voyelle, provoque un hiatus əV, lequel emporte la chute du schwa. La situation est alors analogue à ce que nous avons décrit pour les adjectifs féminins singuliers avec *mal* (*mal(e) heure - malheur, mal(e) aventure - malaventure*, etc., cf. sec. 5.2.3.2, § *Mal + nom*, p. 149).

Encontre. La même collocation avec le pronom personnel « eux » se retrouve pour *encontre*, cette fois-ci de manière presque systématique : six des sept occurrences en défaut de <e> sont suivies du pronom personnel « eux » (6.64 - 6.69), tandis que l'occurrence restante voit la préposition *a* s'interposer entre *contre* et le pronom personnel objet de troisième personne du singulier *lui*, ce qui autorise le défaut de <e> grâce à l'initiale vocalique de *a* (6.70).

- (6.64) *vos alez encontr eus*
 « vous allez à leur rencontre »
 (alia, p. 13, trad. pers.)
- (6.65) *si virent loing venir trotant encontr eus un vallet a pie*
 « et ils virent venir au loin, trottant à leur rencontre, un valet de pied »
 (amad, p. 39, *id.*)
- (6.66) *li venicien virent les grius qui estoient venu seur le rivoage tout armé encontr aus*
 « les Véniciens virent les Grecs qui étaient venus sur le rivage, en armes, à leur rencontre »
 (clari2, p. 49, *id.*)
- (6.67) *en la bataille qu il et sa gent matez rendront toz ceaux qui encontr eaux viendront*
 « en la bataille où lui et ses hommes abbatront tous ceux qui viendront les affronter »
 (eneas, p. 13, trad. pers.)

turp : ANON. (1217*). *Chronique dite Saintongeaise*. Dees : [Vendée, Deux-Sèvres]; compo. saint., ca 1217; ms. poit., ca 1275; MS (MANDACH 1970); DEAF : Turpin¹M.

22: P.ex. GD « contreater », *aller à l'encontre de*; « contrebaillier », *rendre*; « contredire », *défendre, interdire, disputer*; « contrecorre », *courir contre*; etc.

23: P.ex. GD « contredicte », *contradiction, opposition*; « contrecorement », *course en sens opposé, attaque*; « contrefait » et « contrefaiture », *imitation*; etc.

24: P.ex. « contrefaissié » : *garni de fascies*; « contrehastier » : *qui se tient toujours au coin du feu*; etc.

alia : ANON. (1185*). *Roman d'Alexandre*. Ms. A; Dees : Vendée, Deux-Sèvres; compo. pic. et occ. orient., ca 1185; ms. poit., ca 1225; MS (ANON. 1937); DEAF : AlexArsL.

amad : ANON. (1210*). *Amadas et Ydoine*. Ms. P; Dees : Aisne; compo. a.n., ca 1210; ms. pic., ca 1289; CR3 (ANON. 1974); DEAF : AmYdR.

clari2 : ROBERT DE CLARI (1217*). *La conquête de Constantinople*. Dees : Somme, Pas-de-Calais; compo. pic., ca 1217; ms. pic., 1300; MS (ROBERT DE CLARI 1956); DEAF : RobClariL.

eneas : ANON. (1160*). *Éneas*. Ms. A; Dees : Haute-Marne; compo. norm., ca 1160; ms. lorr., ca 1200; CR2 (SALVERDA DE GRAVE 1964-68).

- (6.68) *voit les le roi encontr eus s est levez*
« le roi les voit et s'est levé à leur rencontre »
(nimb1, p. 13, trad. pers.)
- (6.69) *voit les le roy encontr eus s est levez*
« le roi les voit et s'est levé à leur rencontre »
(nimb2, p. 11, *id.*)
- (6.70) *est coruz encontr a lui*
« [il] a couru à sa rencontre »
(martin2, p. 12, *id.*)

Tout comme pour « contre eux », le défaut de <e> dans le syntagme « encontre eux » est minoritaire : il est attesté dans les six occurrences en (6.64 - 6.69), mais le corpus contient également 33 occurrences du même syntagme sans défaut de <e>, dont deux exemples sont donnés en (6.71 - 6.72).

- (6.71) *elles se leverent toutes encontre eulz*
« elles se levèrent toutes à leur rencontre [= quand ils entrèrent] »
(helc, p. 22, trad. pers.)
- (6.72) *ele est venue encontre eus*
« elle est venue à leur rencontre »
(ombre, p. 7, *id.*)

En somme. Il est difficile d'étudier les effets de collocation des prépositions dans la mesure où, par définition, la préposition implique un *type* syntaxique (PRÉP + syntagme nominal). Néanmoins, pour nos trois prépositions se terminant par /trə/, la construction PRÉP + PRO_{PERS-MASC}3pl CR « eux » est spécifiquement associée à des défauts de <e>. Cette association entre défaut de <e> et PRÉP + PRO_{PERS}3MASCpl CR est majoritaire pour « entre » (393 occ. pour 523 occ.), et minoritaire pour pour « contre » (6 occ. en défaut pour 32 occ. du syntagme) et « encontre » (6 occ. en défaut pour 33 occ. du syntagme). De fait, « eux » est le seul pronom personnel à initiale vocalique à apparaître régulièrement après « contre » et « encontre » ; il ne s'agit donc probablement pas d'une spécificité du PRO_{PERS}3pl mais du fait qu'il est le seul pronom à apparaître suffisamment souvent après « entre » et « (en)contre » pour générer des défauts de <e>. À plus faible fréquence, le même effet se retrouve avec le PRO_{PERS}FÉM3pl CR « eles ». Les occurrences *contr aval* et *contr amont* semblent témoigner d'un effet analogue ; ici, fréquence d'association et grammaticalisation viennent plus explicitement mettre à mal la frontière entre les deux mots. Dans notre analyse, ce ne sera donc pas l'effet de la collocation avec « eux » (ou « eles ») que nous retiendrons, mais le fait que les types PRÉP + PRO_{CR} favorisent les élisions.

6.2.4 Distribution contextuelle

La fig. 6.12 en page suivante indique le taux moyen de défaut de <e> de chaque texte, pour les six lemmes retenus, sous la forme de graphiques de séries temporelles concernant le contexte préconsonantique (a, c) ou le contexte prévoicalique (b, d). Elle comporte également un graphique en boîte à moustache (e) illustrant le contraste entre les deux contextes, ainsi que la mesure de la différence de ces deux distributions par le

nimb1 : ANON. (1150*). *Le charroi de Nîmes*. Ms. B1; Dees : Aisne; compo. nil, ca 1150; ms. Paris, ca 1335; MS1 (transcr. inédite L. Schoesler).

nimb2 : ANON. (1150*). *Le charroi de Nîmes*. Ms. B2; Dees : Normandie; compo. nil, ca 1150; ms. Paris, ca 1335; MS1 (transcr. inédite L. Schoesler).

martin2 : PÉAN GÂTINEAU (1125*). *Vie de saint Martin*. Dees : Vendée, Deux-Sèvres; compo. tourr., ca 1125; ms. poit., ca 1290; MS (PÉAN GÂTINEAU 1896); DEAF: PeanGat51.

helc : ANON. (1285*). *Le roman de Helcanus*. Dees : région parisienne; compo. pic., ca 1285; ms. frc, ca 1290; CR2 (NIEDZIELSKI 1966); DEAF: HelcanusN.

ombre : JEAN RENART (1221*). *Le Lai de l'Ombre*. Dees : [Marne]; compo. pic., ca 1221; ms. frc, ca 1288; CR1 (JEAN RENART 1913); DEAF: OmbreB¹.

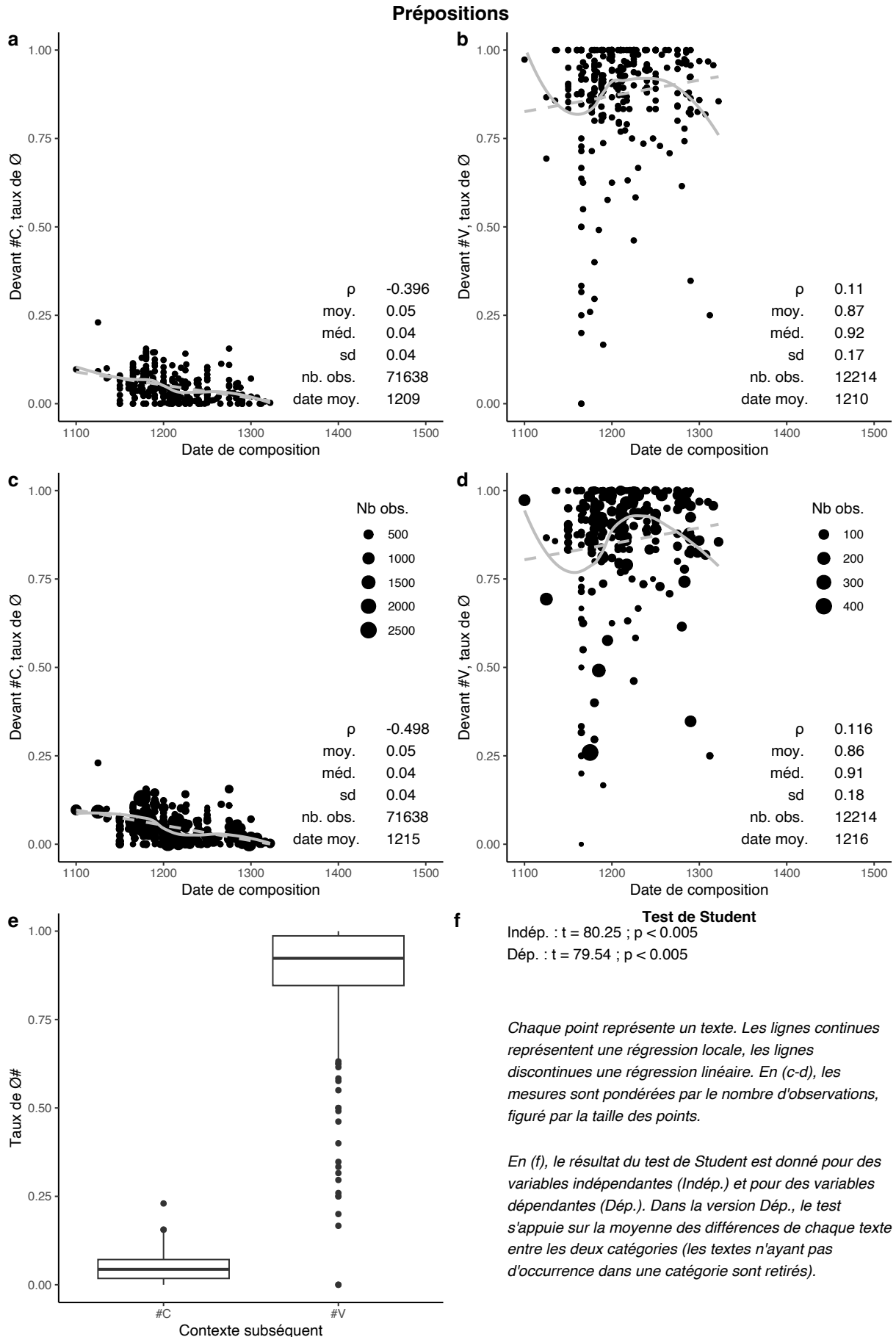


FIGURE 6.12 – Distribution contextuelle des variantes de prépositions en défaut de <e>.

test de Student (f). Comme l'illustre cette figure, sur les six lemmes sélectionnés, la différence entre le contexte prévocalique et le contexte préconsonantique est frappante : le taux moyen de défaut de <e> par texte devant #C est de 5%, tandis qu'il est de 87% ou 86% devant #V, selon que la moyenne n'est pas ou est pondérée par le nombre d'occurrences de chaque texte. Aussi, le contexte prévocalique a un effet marqué sur la génération des défauts de <e> des prépositions ; comme indiqué par le test de Student en fig. 6.12.f, cet effet est très fort ($t = 80,25$) et hautement significatif ($p < 0,005$). Par comparaison, pour les adjectifs féminins singuliers, cet effet était significatif ($p < 0,005$) mais beaucoup moins fort ($t = 4,84$). La comparaison des graphiques de séries temporelles en fig. 6.12.a-d montre deux dynamiques opposées : le taux moyen de défaut de <e> devant initiale consonantique décroît avec le temps ($\rho = -0,4$, voire $-0,5$ après pondération par le nombre d'occ. par texte), tandis qu'il augmente devant initiale vocalique ($\rho = 0,11$, voire $0,12$ après pondération). Par ailleurs, le nombre d'occurrences devant initiale consonantique (71 638) est 6 fois plus élevé que devant initiale vocalique (12 214), ce qui explique que le graphique de série temporel général (non-contextuel) en fig. 6.2 *supra* ressemble bien plus aux graphiques concernant le contexte préconsonantique (fig. 6.12.a,c) qu'à ceux concernant le contexte prévocalique (fig. 6.12.b,d).

Contrairement à la situation des adjectifs féminins singuliers, la distribution des données présente, ici, pour chaque contexte, un degré de cohérence plus élevé. La décroissance des taux de défaut de <e> devant initiale consonantique, représentée par la droite de régression en fig. 6.12.a,c, correspond à une corrélation linéaire relativement marquée ($\rho = -0,4$), particulièrement lorsque les taux sont pondérés par le nombre d'occurrences des textes ($\rho = -0,5$). Dans les deux cas, la corrélation, bien que modérée, est hautement significative ($p < 0,005$).²⁵ C'est moins le cas devant initiale vocalique : la corrélation est faible ($\rho = 0,11$, et $0,12$ dans la version pondérée) et n'est significative que dans sa version où les taux sont pondérés par le nombre d'occurrences des textes ($p = 0,048$), alors que dans la version non pondérée la valeur- p est supérieure au seuil de $0,05$ ($p = 0,06$). Aussi, alors que le jeu de données devant initiale consonantique est assez cohérent, il l'est moins devant initiale vocalique, ce qui se traduit également par un écart-type quatre fois plus élevé devant initiale vocalique ($\sigma = 0,17$ ou $0,18$) que devant initiale consonantique ($\sigma = 0,04$). Visuellement, le contexte préconsonantique dispose de très peu d'*outliers* et les valeurs de ceux-ci sont moins extrêmes, tandis que le contexte prévocalique dispose de nombreux points dont les valeurs sont, vis-à-vis de la droite de régression, anormalement basses.

Enfin, en lien avec la faible corrélation linéaire des données prévocaliques, on observe sur les graphiques de série temporelle du contexte prévocalique (fig. 6.12.b,d) un pic de concentration des plus hautes valeurs de taux moyens de défaut de <e>, quelque peu analogue à celui observé pour les adjectifs féminins singuliers. La régression locale sur ces figures permet de visualiser ce pic, tel que mise en avant par la figure marginale 6.13. Cet effet est quelque peu décalé dans le temps ; pour les adjectifs féminins singuliers, il culminait dans la seconde moitié du XII^e s., alors qu'ici le taux moyen de défaut de <e> par texte devant initiale vocalique atteint un plateau seulement à la fin du XII^e s. et se maintient sur la première moitié du XIII^e s. Pour les adjectifs féminins singuliers, nous

25: La valeur- p de la corrélation n'est pas inscrite sur le graphique, et nous la reproduisons ici depuis la console R.

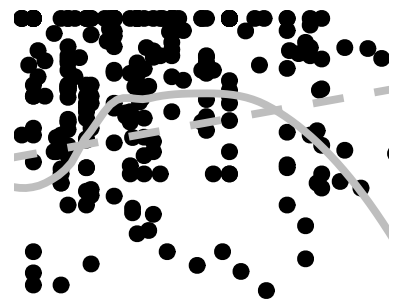


FIGURE 6.13 – Extrait de la fig. 6.12.b, montrant le plateau de la régression locale autour de 90% entre 1200 et 1250.

avons expliqué ce pic par une concentration particulièrement forte de textes anglo-normands dans la seconde moitié du XII^e s., lesquels étaient plus générateurs de défaut de <e> que les textes continentaux; cette explication ne peut donc pas fonctionner pour les prépositions, et nous en proposerons une autre *infra* (sec. 6.3.5).

Étant donné que nous avons noté que l'effet du contexte était fortement variable en fonction du lemme, il convient ici d'observer la situation des prépositions à l'exception de « sor » et « desor », puis d'observer la situation de ces deux seules prépositions.

6.2.4.1 « De », « entre », « contre » et « encontre »

La fig. 6.14 en page suivante reproduit la même présentation des données, mais uniquement pour les lemmes « de », « entre », « contre » et « encontre ». Elle montre que la situation devant initiale vocalique de ces quatre lemmes est très proche de celle de l'ensemble des six lemmes. Le taux moyen de défaut de <e> par texte reste identique (88%, ou 87% après pondération; fig. 6.14.b,d). En revanche, devant initiale consonantique (fig.6.14.a-c), le taux moyen de défaut de <e> est réduit à zéro, avec uniquement quelques textes disposant de taux positifs. Comme le confirmeront les données de la sous-section suivante, dédiée à « sor » et « desor », ce sont donc ces deux lemmes qui sont reponsables d'une écrasante majorité des défauts de <e> devant initiale consonantique observés *supra*. Sur les lemmes « de », « contre » et « encontre », l'effet du contexte prévoicalique mesuré par le test de Student (fig.6.14.f) est donc plus marqué que sur l'ensemble des lemmes ($t = 85,77$; $p < 0,005$).

Devant initiale vocalique, la dispersion des données demeure élevée ($\sigma = 0,17$) et le retrait des 158 occurrences en défaut de <e> de « sor » et « desor » ne change guère l'ajustement des données au modèle de la régression linéaire. La valeur des coefficients de corrélations reste identique et donc très faible ($\rho = 0,11$, et $0,12$ dans la version pondérée), et seule la valeur- p du jeu de données pondérées diminue, indiquant que l'alignement des données sur le modèle linéaire est légèrement plus significatif ($p = 0,045$). Néanmoins, cette situation, qui mélange « de », « entre », « contre » et « encontre », est surtout représentative du lemme « de » (9978 occ. défectueuses sur 77 369 occ.), du fait du faible poids des autres lemmes. Il est donc nécessaire d'observer ce qu'il en est pour ces lemmes en dissociant « de », « entre » et « contre » et « encontre ».

« De ». Du fait de la très forte prédominance de « de » sur les autres lemmes, la figure synoptique concernant uniquement le lemme « de » est pratiquement semblable à celle concernant les trois lemmes; elle est rejetée en annexes (fig. E.5, p. 614). Ce qu'il faut observer, ici, ce sont les quelques cas de défaut de <e> devant initiale consonantique, ainsi que les cas d'absence de défaut de <e> devant initiale vocalique.

Concernant les cas de défaut de <e> devant initiale consonantique, il convient de noter à quel point ils sont marginaux : ils concernent 22 occurrences sur 66 078 occurrences préconsonantiques, soit 0,03%. Ces cas nous apparaissent tous douteux, sans qu'il n'ait été possible, à l'étape d'exclusion des formes, d'objectiver un motif d'exclusion. Un examen plus approfondi révèle que, pour au moins une partie des ces formes,

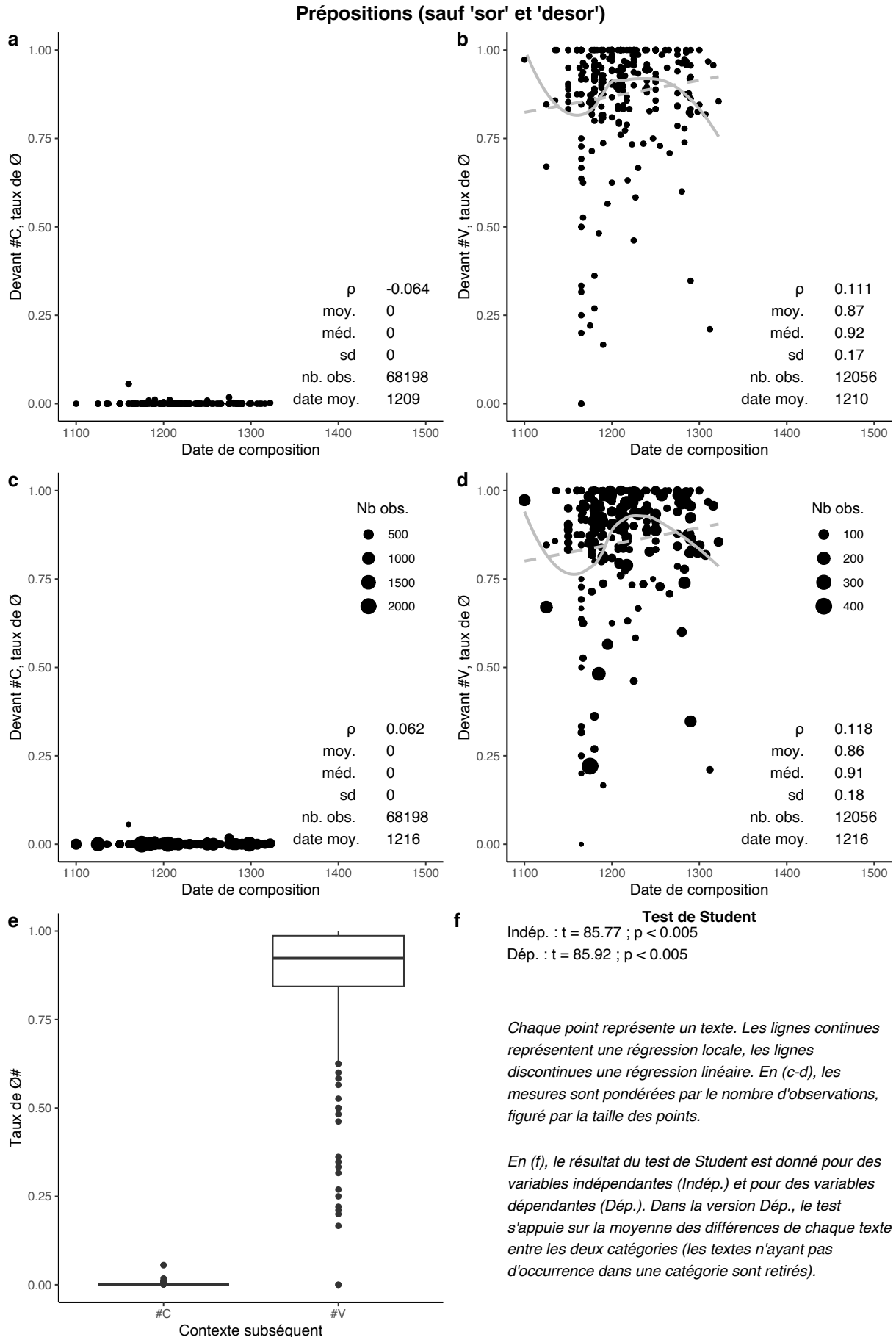


FIGURE 6.14 – Distribution contextuelle des variantes de prépositions en défaut de <e>, à l'exception de « sor » et « desor ».

l'édition du NCA est corrompue. Ainsi, en (6.73), le v. 1007 comporte, tel qu'édité par le NCA, le syntagme *d sarrazins* pour *de sarrazins*. Or non seulement d'autres éditeurs obtiennent *de sarrazins* et non *d sarrazins*, mais même l'édition qu'a utilisée Dees (SEGRE 1971) comporte bien *de sarrazins*, ce qui est conforme au ms., reproduit en fig. 6.15.

- (6.73) NCA : *dist oliver 'sire cumpainz ce crei :*
d sarrazins purum bataille aveir
 BFM : *Dist Oliver : « Sire cumpainz, ce crei,*
De Sarrazins purum bataille aveir. » (id. BÉDIER, SEGRE)²⁶
 Dipl. *D ift oluer fire cūpainz ce crei.*
d e farrazins purū bataille auer.
 « Olivier dit : « Sire compaignon, je crois
 que nous pourons avoir bataille avec les Sarrasins. » »
 (Roland : rolandox (NCA), roland (BFM), v. 1006-1007, trad.
 SHORT 1990)

26: Le signalement du discours direct est différent chez SEGRE (1971, p. 190), mais la leçon est identique.

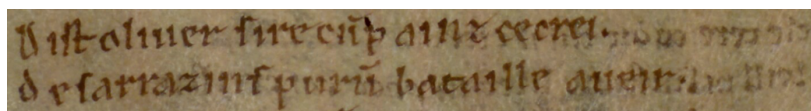


FIGURE 6.15 – *Roland*, ms. d'Oxford, fol. 18v, l. 24-25 (v. 1006-1007). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

La situation est analogue pour la seconde occurrence de *d* devant initiale consonantique de *Roland*. Au v. 3765 (6.74), le NCA dispose de *d france* pour *de france*, alors que les éditeurs, Segre compris, éditent *de france* et que *de* se lit sans difficulté sur le ms., tel que reproduit en fig. 6.16.

- (6.74) NCA : *veit cels d france e tuz les juteurs*
 BFM : *Veit cels de France e tuz les jugeurs,* (id. BÉDIER, SEGRE)
 Dipl. *V eit cels de france 7 tuz les iuteurs.*
 « Il regarde bien ceux de France et tous ses juges, »
 (Roland : rolandox (NCA), roland (BFM), v. 3765, trad. SHORT
 1990)

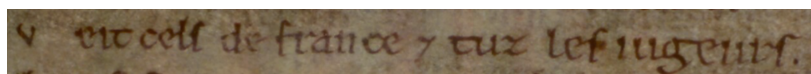


FIGURE 6.16 – *Roland*, ms. d'Oxford, fol. 68r, l. 9 (v. 3765). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

Nous ne savons pas d'où proviennent ces erreurs dans le NCA. Dans la mesure où Segre n'indique pas d'intervention éditoriale et où le ms. est sans ambiguïté pour ces deux occurrences, ces incongruités du NCA ne sont pas du même type que celle relevée en (6.3), où la lacune dans *paienviat* (= *paien*<*s vos en en*>*viat*) pouvait expliquer la lacune du NCA. Par ailleurs, si, dans certaines occurrences, *d* peut former avec la consonne initiale subséquente une attaque potentiellement licite (/dr/ en 6.75, /ds/ en 6.76), ce n'est pas le cas dans la plupart des occurrences (p.ex. */db/ en 6.77 - 6.78).

- (6.75) *vestus d riches dras*

cambrai : ANON. (1190*). *Raoul de Cambrai*. Dees : Ardennes ; compo. Nord-Est, ca 1190 ; ms. pic., ca 1225 ; CR1 (MEYER et LONGNON 1882) ; DEAF : RCambrM.

- « vêtu de riches vêtements »
(cambrai, p. 28, trad. pers.)
- (6.76) *arnauz de bellande passa premiers les pors d cysaire*
« Arnaud de Bellande passa d'abord par les ports de Cysaire »
(pseuturp, p. 14, *id.*)
- (6.77) *ne t anmera ne d bien ne te fera preudon*
« [elle] ne t'aimera pas, ni, d'une bonne façon, ne fera de toi un homme preux »
(contre, p. 34, *id.*)
- (6.78) *ja n i perdrais le montant d balois*
« jamais je n'y perdrais le montant d'un rubis »
(cambrai, p. 12, *id.*)

pseuturp : ANON. (1290*). *The burgundian translation of the Pseudo-Turpin Chronicle*. Dees : Haute-Saône; compo. bourg., ca 1290; ms. *idem*; MS (WALPOLE 1949, 1948); DEAF : Turpin⁷W.

contre : ANON. (1322*). *Le Roman de Renart le contrefait, fin de la 6^e et dernière branche*. Dees : Aube; compo. champ. mérid., ca 1322; ms. nil, ca 1337; CR2 (LEMAÎTRE et RAYNAUD 1914); DEAF : RenContr¹R.

Au demeurant, ces occurrences de *d* devant initiale consonantique sont extrêmement rares, ne représentant que 0,03% des occurrences devant initiale consonantique. Dans la mesure où nous avons pu démontrer que certaines sont éditorialement fausses (cf. 6.73, 6.74) et où d'autres représentent des configurations phonologiques qui rendent l'occurrence suspecte (du moins dans la mesure où elle représenterait non une abréviation mais un trait phonologique), il convient d'ignorer ces occurrences, en les rangeant dans le bruit de fond inévitable des données médiévales. Il est intéressant, ici, de noter que ce bruit de fond ne provient pas uniquement des ms. eux-mêmes ou des éditions, mais également du corpus en lui-même, en ce qu'il semble s'écarter parfois de son édition de référence.

« **Entre.** » La fig. 6.17 en page suivante reproduit la présentation des données vue ci-dessus, mais uniquement pour le lemme « entre ». La différence entre les initiales vocaliques et consonantiques y est patente, avec aucune occurrence *entr* devant initiale consonantique et un taux moyen d'occurrence du type *entr* de 68% devant initiale vocalique. Aucun effet diachronique ne semble ressortir des graphiques de séries temporelles en fig. 6.17.b,c; la concentration de textes dans la seconde moitié du XII^e s. s'effectue sur toute la hauteur de l'échelle, et l'on trouve des textes dotés d'aucune occurrence défectueuse sur toute la période, comme on trouve des textes dotés uniquement d'occurrences défectueuses sur toute la période. Enfin, le test de Student en fig. 6.17.f permet de mesurer la puissance de l'effet du contexte prévocalique; celui-ci est inférieur à l'effet rencontré pour la préposition « de » (« de » : $t = 85,77$: « entre » : $t = 23,49$), ce qui s'explique par la grande quantité de textes disposant de taux moyens et de textes ne disposant d'aucune occurrence défectueuse.

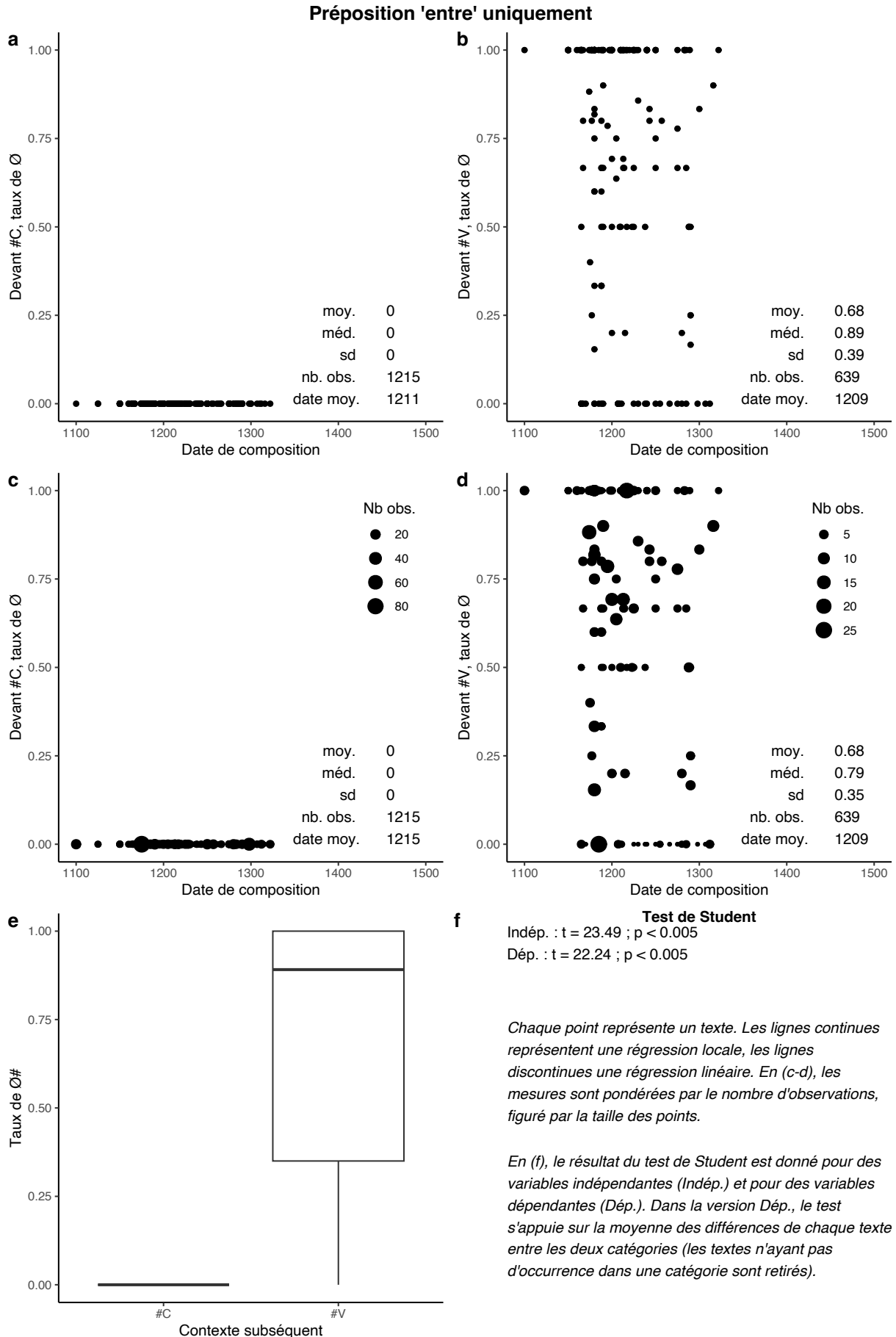


FIGURE 6.17 – Distribution contextuelle des variantes de prépositions en défaut de <e>, lemme « entre » uniquement.

« **Contre** » et « **encontre** ». La fig. 6.18 en page suivante reproduit la même présentation des données, mais pour les lemmes « contre » et « encontre ». Comme exposé *supra*, les formes en défaut de <e> de « contre » et « encontre » se rencontrent dès la moitié du XII^e s., et n'apparaissent que devant initiale vocalique. Parallèlement, sur toute la période, on trouve de nombreuses formes sans défaut de <e>, tant devant initiale consonantique que devant initiale vocalique, comme représenté en fig. 6.18.a-d. Un effet diachronique est peut-être en jeu, dans la mesure où les textes ne comportant que des occurrences (*en*)*contr* ou comportant des taux mixtes d'occurrences (*en*)*contr* et (*en*)*contre* se regroupent dans la seconde moitié du XII^e s. et dans la première moitié du XIII^e s. C'est certes la période la plus fournie de notre corpus, mais la raréfaction drastique des formes défectueuses à partir du milieu du XIII^e s., tandis que les formes non défectueuses continuent à être attestées, ne permet pas de conclure à une simple influence de la composition du corpus.

Étant donné que ces lemmes sont relativement rares devant initiale vocalique (126 observations sur l'ensemble du corpus), soit moins de une par texte, rares sont les textes à disposer de plusieurs observations. De fait, si le test de Student en fig. 6.18.f mesure bien un effet du contexte prévocalique sur le taux de défaut de <e>, cet effet est assez faible ($t = 4,62$) tout en étant significatif ($p < 0,005$). La force de l'effet, équivalente à celle mesurée pour les adjectifs féminins singuliers ($t = 4,84$), témoigne ici d'une simple condition nécessaire : on ne peut trouver de défaut de <e> sur « contre » et « encontre » que devant initiale vocalique. Contrairement à ce que nous avons pu observer pour « de » ou « entre », cette condition est loin d'être suffisante, et les taux de défaut devant initiale vocalique restent faibles (24%, ou 16% après pondération par le nombre d'occurrences des textes).

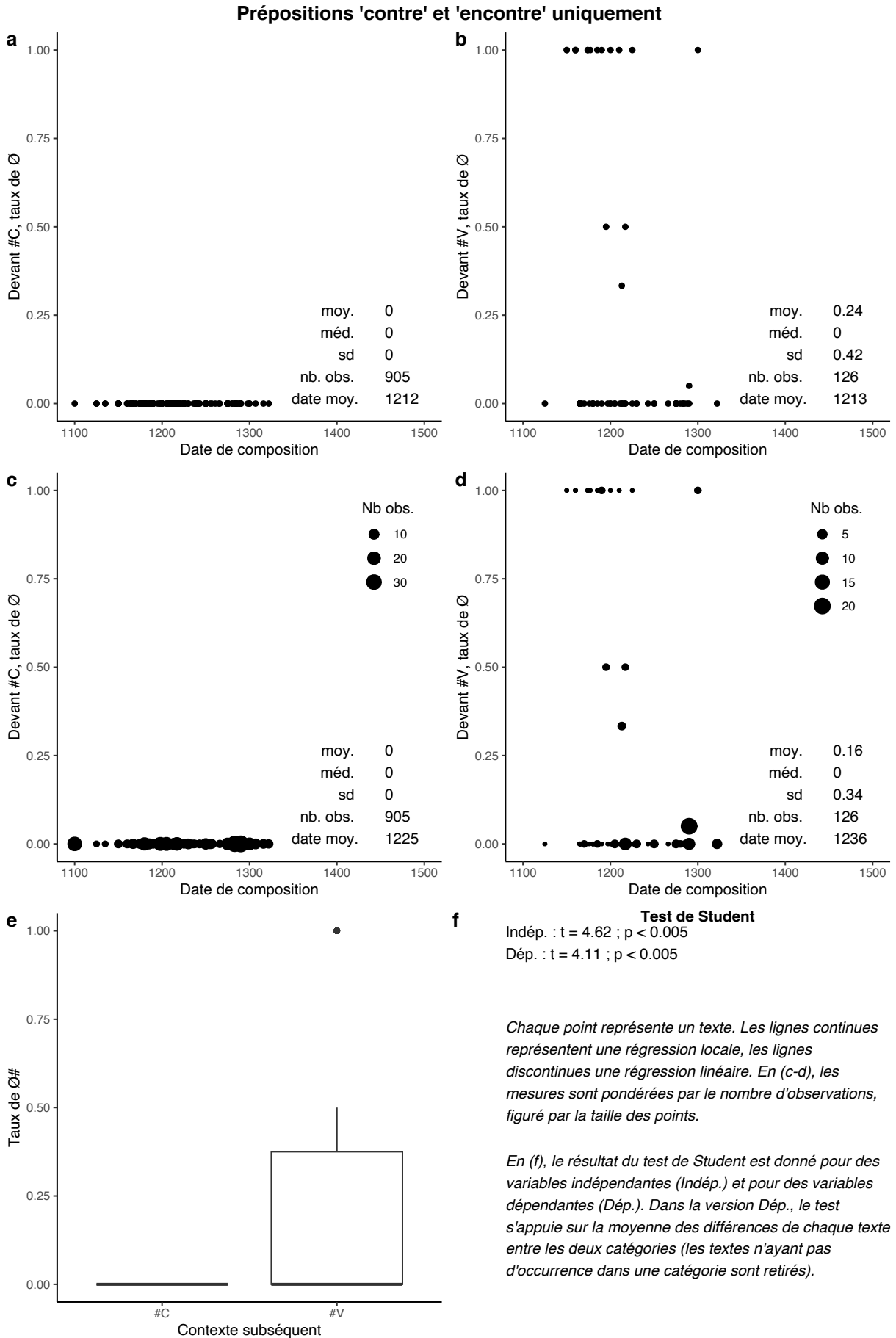


FIGURE 6.18 – Distribution contextuelle des variantes de prépositions en défaut de <e>, lemmes « contre » et « encontre ».

6.2.4.2 « Sor » et « desor »

La fig. 6.19 en page suivante reproduit toujours la même présentation des données, mais pour les lemmes « sor » et « desor ». La situation, ici, est radicalement différente : l'effet du contexte est nul ($t = 0,02$) et non significatif ($p = 0,981$). Dans la version à variables dépendantes du test de Student — c'est-à-dire portant uniquement sur les textes disposant d'occurrences dans les deux contextes, un léger effet se dégage, cette fois-ci en défaveur du contexte prévocalique ($t = -1,72$), mais celui-ci, bien que plus significatif que le test à variables non dépendantes, n'est pas significatif non plus ($p = 0,089$) ; cette différence signifie simplement qu'on trouve en moyenne plus de défauts de <e> devant initiale consonantique que devant initiale vocalique dans les textes contenant des occurrences dans les deux contextes, mais que la probabilité que ce résultat soit dû au hasard est forte. Dans les deux contextes, les taux moyens de défaut de <e> par texte sont très élevés (moy. = 0,95 ou 0,97% __#C et 0,95 ou 0,96% __#V) ; ce sont les taux inférieurs à cette moyenne qui se distinguent comme des *outliers* vis-à-vis d'une concentration de données très marquée autour de 100% de défaut de <e>. En effet, comme indiqué par le graphique en boîte à moustache en 6.19.e, plus de 75% des textes ont un taux de défaut de <e> de 100% (les trois derniers quartiles sont écrasés à hauteur de 100%).

Il faut néanmoins rappeler que ces graphiques et mesures ne traitent pas, dans l'absolu, des prépositions « sor » et « desor », mais des occurrences de ces deux lemmes qui sont annotées comme étant des prépositions (pos="PRE"). Or nous avons vu *supra* (p. 197) que l'annotation en parties du discours de ces deux lemmes était parfois fautive, et nous n'avons pas été en mesure d'établir son degré de justesse. Aussi, il est possible que parmi nos occurrences, certaines ne soient pas des prépositions, et que des attestations prépositionnelles manquent dans nos données, n'étant pas annotées correctement dans le corpus (pos!="PRE").

Si l'on accepte ces données, malgré l'incertitude de leur annotation morphosyntaxique, il convient de conclure que les défauts de <e> des prépositions « sor » et « desor » sont insensibles au contexte subséquent, et que ces défauts dominent largement. Diachroniquement, comme avancé *supra*, les formes sans <e> dominent sur l'ensemble de la période, ce qui témoigne d'un processus radicalement différent de celui qui affecte nos autres prépositions. Dans notre analyse linguistique, nous considérerons donc que *sor* et *desor* représentent un développement idiosyncratique, qui doit être isolé de celui des autres prépositions pour ne pas en polluer les résultats. Ce sont les résultats des autres prépositions qui sont importantes pour l'établissement de l'éliision.

6.2.4.3 En somme

Pris ensemble, nos six lemmes prépositionnels montrent un effet fort et significatif du contexte : devant initiale vocalique, les taux de défaut de <e> par texte se concentrent entre approximativement 85 et 100%, alors que devant initiale consonantique, ils se concentrent entre approximativement 20% et 0%. On observe également une double dynamique diachronique, le taux moyen de défaut de <e> devant initiale vocalique

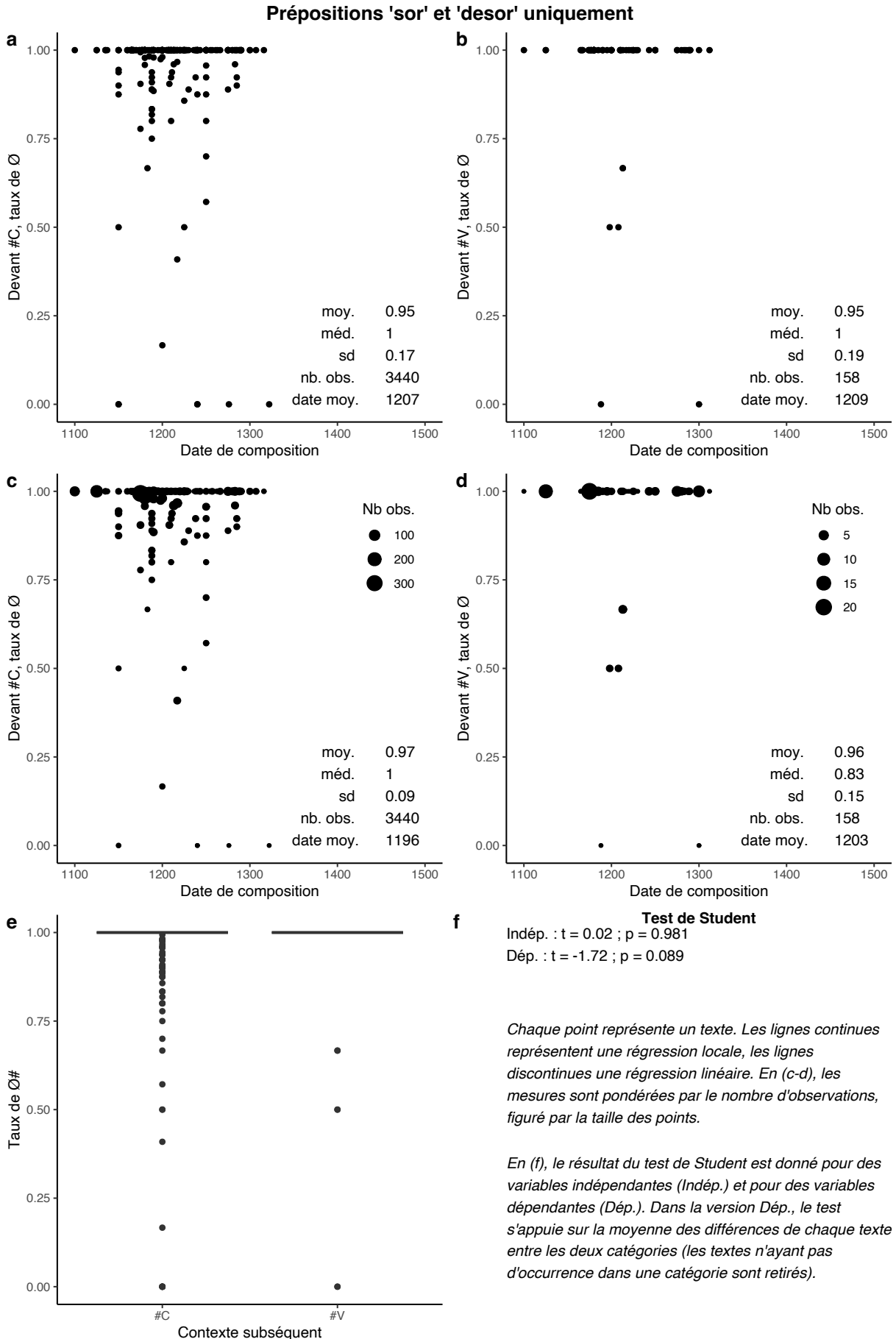


FIGURE 6.19 – Distribution contextuelle des variantes de prépositions « sor » et « desor » en défaut de <e>.

allant croissant alors que les taux de défaut de <e> devant initiale consonantique décroissent. Néanmoins, ces données regroupent deux jeux de données très différents. En isolant les lemmes « de », « entre », « contre » et « encontre », on observe que le défaut de <e> est rarissime voire suspect devant initiale consonantique, alors que, devant initiale vocalique, il demeure à des taux très élevés pour « de » et « entre », et peu élevés pour « contre » et « encontre ». En revanche, les lemmes « sor » et « desor » montrent des taux de défaut de <e> très élevés et stables dans le temps, peu importe le contexte.

6.2.5 Distribution diatopique

Étant donnés les éléments avancés *supra*, la distribution diatopique des formes en défaut de <e> de l'ensemble de nos lemmes est marquée par un bruit de fond assez bas (moy. = 4,35%) et plutôt homogène (min. = 1%; max. = 7%; $\sigma = 2$) devant initiale consonantique, comme représenté en fig. 6.20. Ce bruit de fond correspond aux prépositions « sor » et « desor », dont le poids est contrebalancé par les prépositions « de », qui sont 19 fois plus nombreuses.

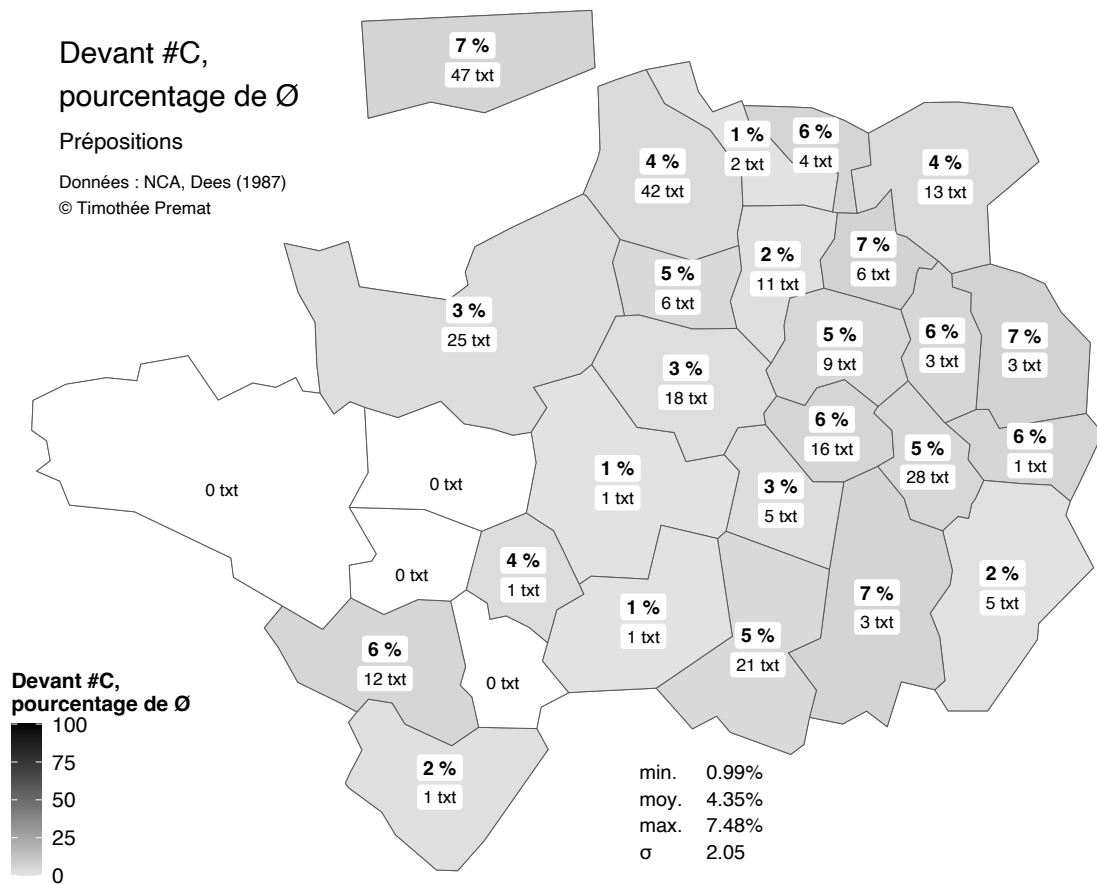


FIGURE 6.20 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, prépositions suivies d'une initiale consonantique.

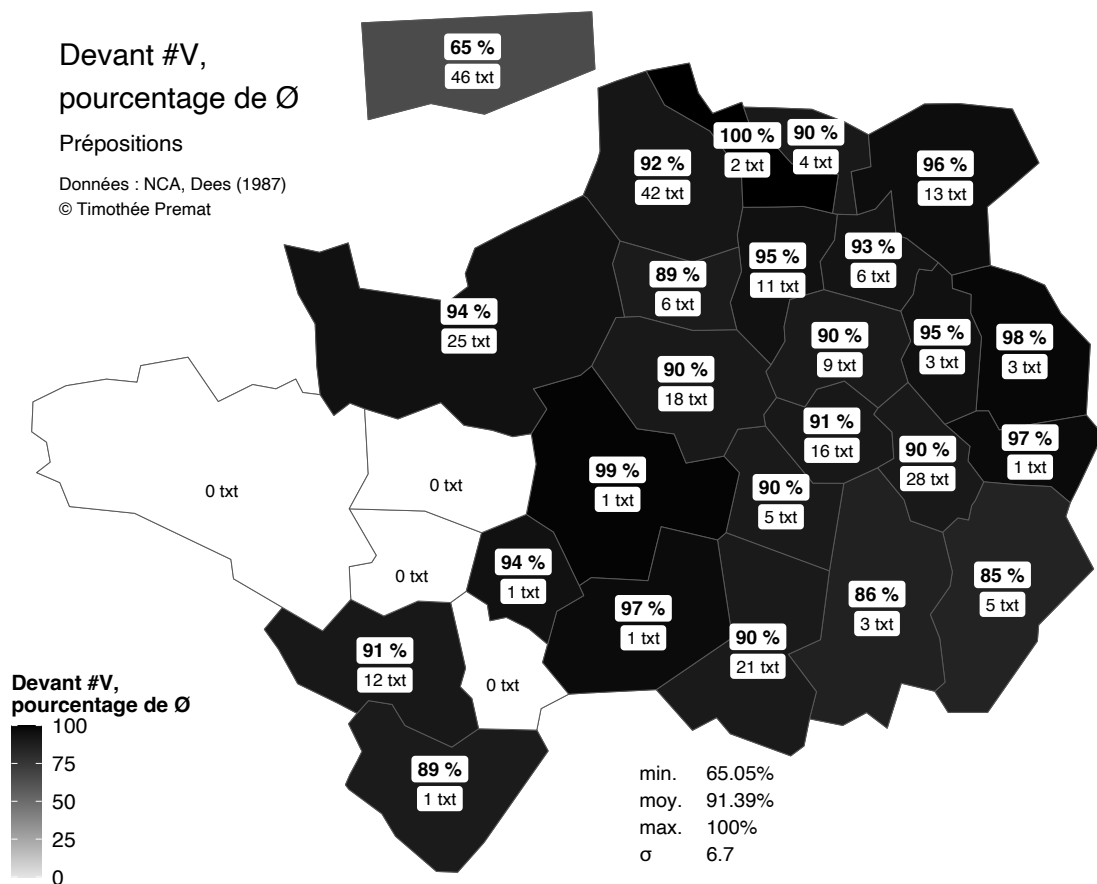


FIGURE 6.21 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, prépositions suivies d'une initiale vocalique.

Devant initiale vocalique, le domaine continental présente des taux de défaut de <e> également très élevés et homogènes, de 85 à 100%, comme représenté en fig. 6.21. Ici, l'effet de la disparition généralisée du <e> de « sor » et « desor » se combine avec la tendance presque systématique au défaut de <e> de « de » et de « entre », tandis que « contre » et « encontre », dont le taux de défaut est bien moins élevé (moy. = 24%), n'ont que peu d'incidence étant donné leur quantité négligeable (126 occ., contre 11 291 pour « de »). En revanche, l'Angleterre se détache du domaine continental par un taux de défaut de <e> moindre, à 65%. Cette particularité anglo-normande sera reprise *infra* (sec. 6.3.5), au stade de la montée en généralisation.

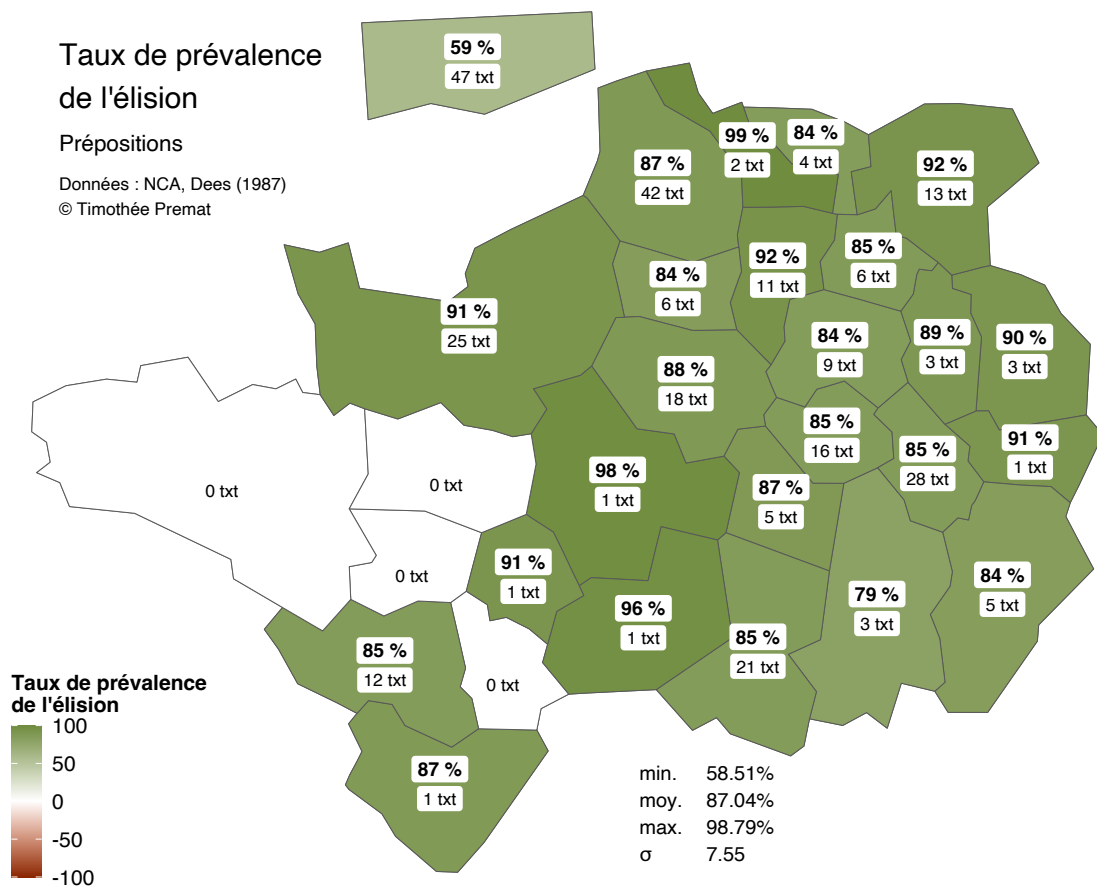


FIGURE 6.22 – Distribution diatopique de la prévalence de l'élision, sur l'ensemble de la période considérée, prépositions.

Du fait de cette différence très marquée entre les deux contextes, qui tient essentiellement à la sensibilité au contexte de la préposition « de » et, dans une moindre mesure, de la préposition « entre », le taux de prévalence de l'élision (fig. 6.22), c'est-à-dire le rapport entre les taux de défaut devant initiale vocalique et ceux devant initiale consonantique, est positif et très élevé sur l'ensemble du domaine continental (de 87 à 99%). Ce taux est en revanche plus faible en Angleterre (59%), par la combinaison d'un taux de défaut devant initiale consonantique élevé (7%) et d'un taux de défaut devant initiale vocalique anormalement bas (65%).

Dans ce qui suit, nous détaillons la distribution diatopique pour « de » (sec. 6.2.5.1), « entre » (sec. 6.2.5.2), « sor » et « desor » (sec. 6.2.5.3. Nous n'aborderons pas la diatopie de « contre » et « encontre » étant donné le trop faible nombre d'occurrences défectueuses (13 et 7 respectivement ; cf. table 6.2 *supra*).

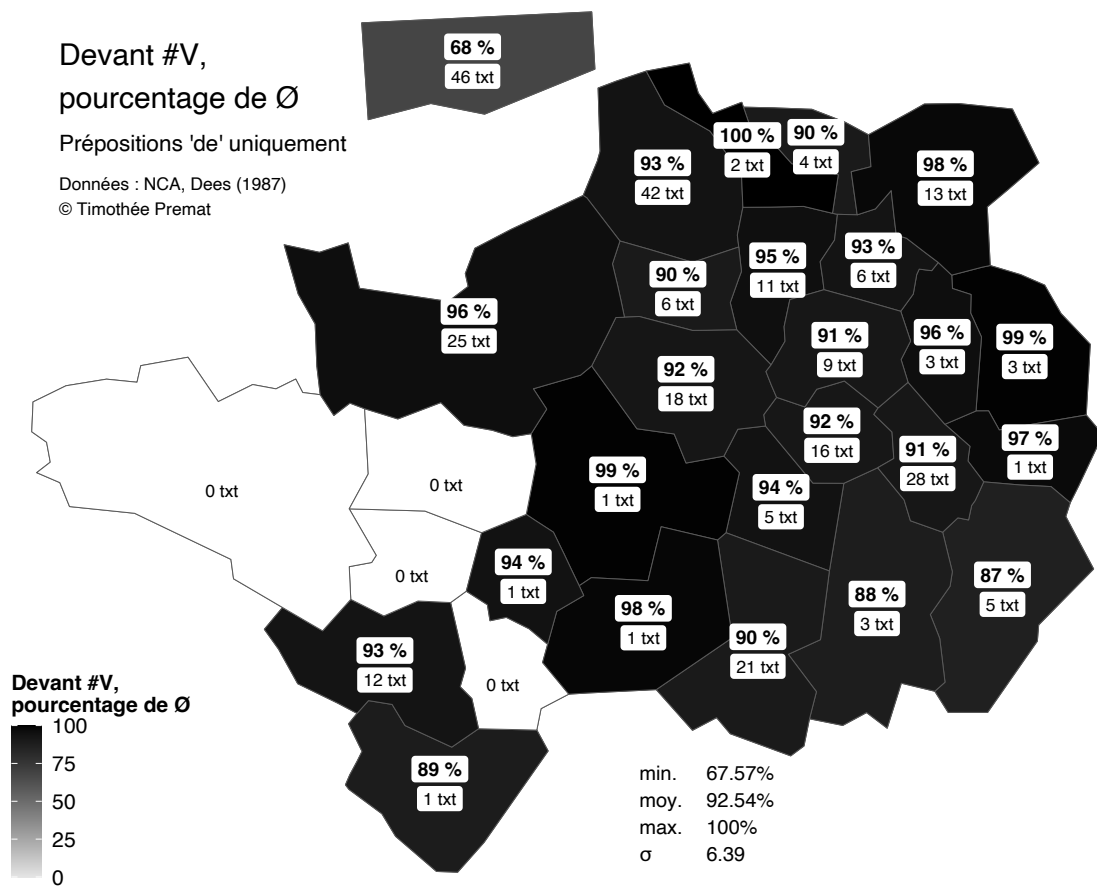


FIGURE 6.23 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, prépositions « de » suivies d'une initiale vocalique.

6.2.5.1 « De »

La carte en fig. 6.23 représente les taux régionaux de défaut de <e> pour le lemme « de » devant initiale vocalique. Dans la mesure où ce lemme ne connaît de défauts de <e> devant initiale consonantique que de manière extrêmement marginale, et dans des occurrences parfois suspectes, nous ne reproduisons pas la carte correspondant à ce contexte. Devant initiale vocalique, la disparition du <e> de ce lemme ressemble à celle de l'ensemble de nos lemmes : le domaine continental dispose de taux élevés, entre 87% et 100%, tandis que l'Angleterre dispose d'un taux moindre, à 68%. Au sein du domaine continental, il est difficile de noter une répartition géographique significative, si ce n'est que la marge méridionale, à l'exception de la région Indre, Cher, dispose de taux de défaut de <e> parmi les moins élevés (87% en Franche-Comté, 88% en Bourgogne, 89% en Charente-Maritime et 90% dans la région Nièvre, Allier); aucune autre région continentale ne dispose de taux inférieurs à 90% et la moyenne des régions s'établit à 92,5%. Les autres marges du domaine continental semblent démontrer une tendance légèrement plus forte au défaut de <e>, à l'exception du Hainaut (90%). En effet,

hormis la marge méridionale déjà mentionnée et le Hainaut, on ne trouve de taux inférieurs à la moyenne qu'entre l'Oise et la Haute-Marne. Si l'on regroupe la spécificité de la marge méridionale et celle, moins marquée, d'une partie du centre et du centre-Est du domaine, les taux les plus faibles se concentrent sur une large bande allant du Centre du domaine (Oise, région parisienne) au Sud-Est, et seules deux régions à taux inférieurs à la moyenne n'entrent pas dans cette distribution : la Charente-Maritime et le Nord.

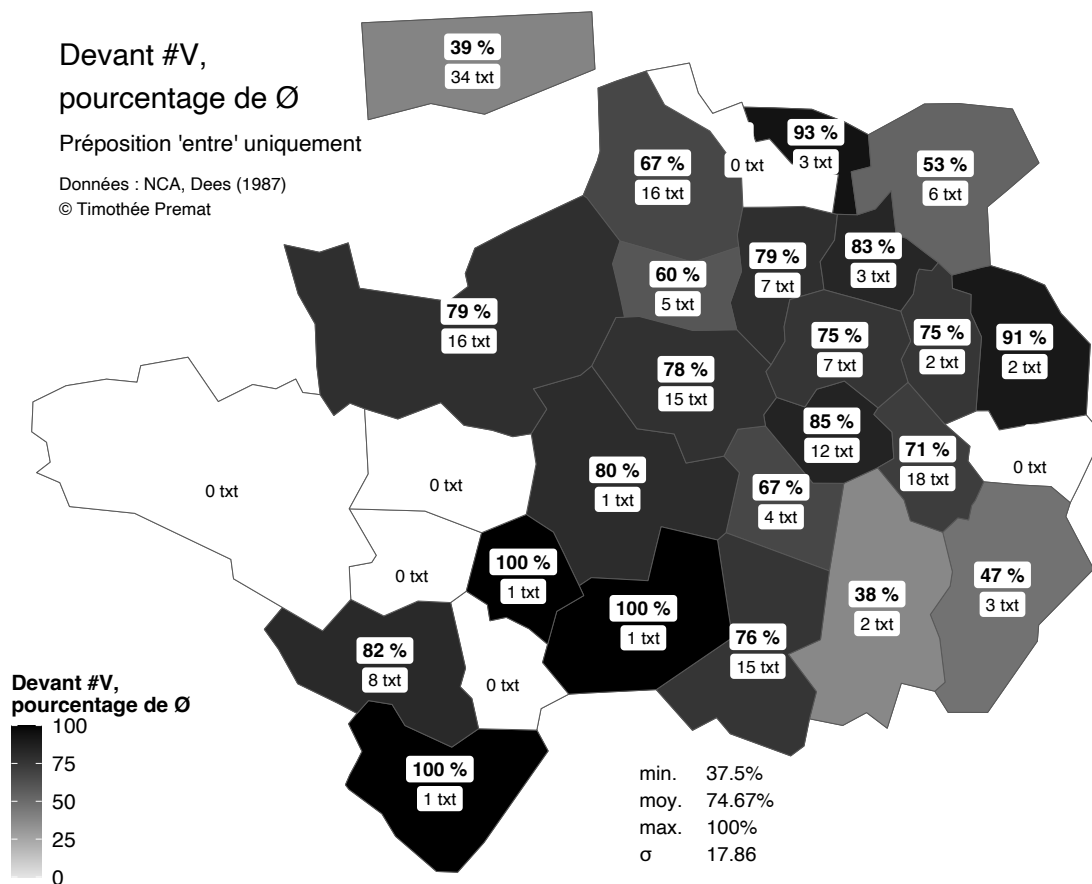


FIGURE 6.24 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, prépositions « entre » suivies d'une initiale vocalique.

6.2.5.2 « Entre »

La carte en fig. 6.24 représente les taux régionaux de défaut de <e> pour le lemme « entre » devant initiale vocalique. L'ensemble du domaine est hétérogène ($\sigma = 17,86$), avec un taux régional moyen (75%) bien inférieur à la moyenne des textes sans considération de la diatopie (89%), ce qui signifie que certaines régions, probablement dotées de peu de textes, tirent vers le bas la moyenne des taux régionaux. Si l'Angleterre se démarque toujours par un taux moyen de défaut de <e> bas (39%), d'autres régions du continent disposent de taux s'écartant significativement de la moyenne des régions (moy. = 74.67%), avec un taux minimal en Bourgogne (38%) et des taux inférieurs à 70% en Franche-Comté (47%), Wallonie (53%), dans l'Oise (60%), l'Yonne et la région Somme, Pas-de-Calais (67%). Ces taux bas ne sont pas systématiquement associés à un nombre de textes faible : c'est le cas pour la Bourgogne (2 textes), la Franche-Comté (3 textes) et l'Yonne (4 textes), mais les autres régions à taux inférieurs à 70% disposent de 5 textes ou plus : l'Oise (5 textes), la Wallonie (6 textes), la région Somme, Pas-de-Calais (16 textes) et l'Angleterre (34 textes). Au contraire, ce sont les taux de 100% qui sont systématiquement associés à

la présence d'un seul texte comportant le lemme prépositionnel « entre », en Charente-Maritime, dans l'Indre-et-Loire et dans la région Indre, Cher. Aussi, ce sont plus les taux de 100% que les taux de moins de 70% qui sont suspects du point de vue de leur assise quantitative.

Géographiquement, les taux les plus faibles n'apparaissent que dans une large moitié Nord-Est du domaine, au Nord d'une diagonale allant de l'Angleterre (incluse) à la Bourgogne (incluse). Cette partie du domaine est néanmoins hétérogène, comportant également de nombreuses régions dont les taux sont supérieurs à la moyenne. Par ailleurs, au sein de cette large moitié Nord-Est, les régions dont les taux sont inférieurs à 70% ne forment pas un ensemble continu, mais se regroupent en deux ensembles continus et une région isolée : l'Yonne, la Bourgogne et la Franche-Comté peuvent être considérées comme formant un ensemble, de même que l'Oise, la région Somme, Pas-de-Calais et l'Angleterre, tandis que la Wallonie est frontalière de régions dont les taux sont supérieurs ou égaux à la moyenne. On ne voit pas clairement se dessiner de zones de transition autour de ces régions continues, mais le grain géographique issu des régions de Dees est probablement trop grossier pour permettre cela.

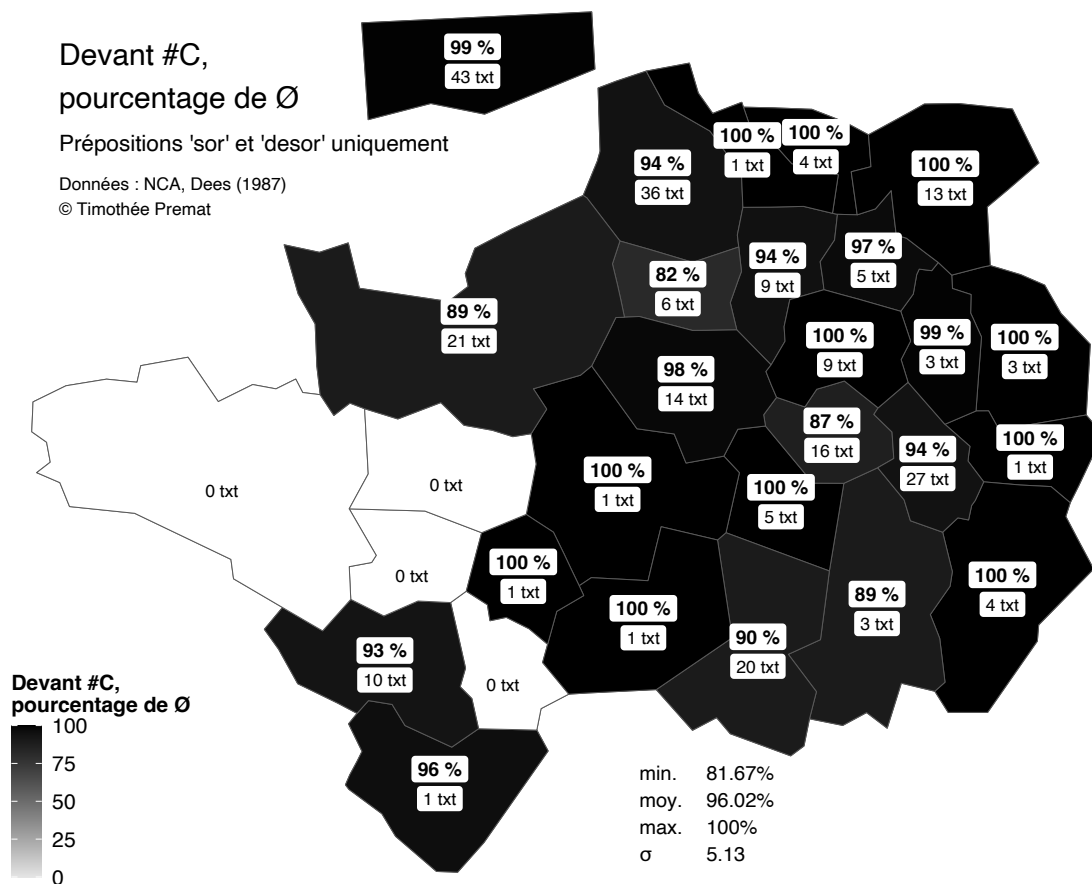


FIGURE 6.25 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, prépositions « sor » et « desor » suivies d'une initiale consonnantique.

6.2.5.3 « Sor » et « desor »

Pour les lemmes « sor » et « desor », il est nécessaire de reprendre les cartes correspondant à chaque contexte, puisque les défauts de <e> apparaissent aussi bien devant initiale vocalique que devant initiale consonnantique. La carte en fig. 6.25 représente les taux régionaux de défaut de <e> de ces lemmes devant initiale consonnantique, et la carte en fig. 6.26, en page suivante, représente ces mêmes taux devant initiale vocalique. Devant initiale consonnantique (fig. 6.23), la situation est assez homogène : toutes les régions disposent de taux de défaut de <e> allant de 82 à 100%, et l'écart-type est réduit ($\sigma = 5,13$). Contrairement à ce que nous avons observé pour les autres lemmes, l'Angleterre adopte l'un des taux les plus élevés (99%), et ne se distingue pas en cela de nombreuses régions continentales : l'Indre-et-Loire, la région Indre, Cher, l'Orléanais, l'Yonne, la Marne, le Nord, le Hainaut, la Wallonie, la Meuse, la région Moselle, Meurthe-et-Moselle, les Vosges et la Franche-Comté disposent de taux supérieurs ou égaux à 99%.

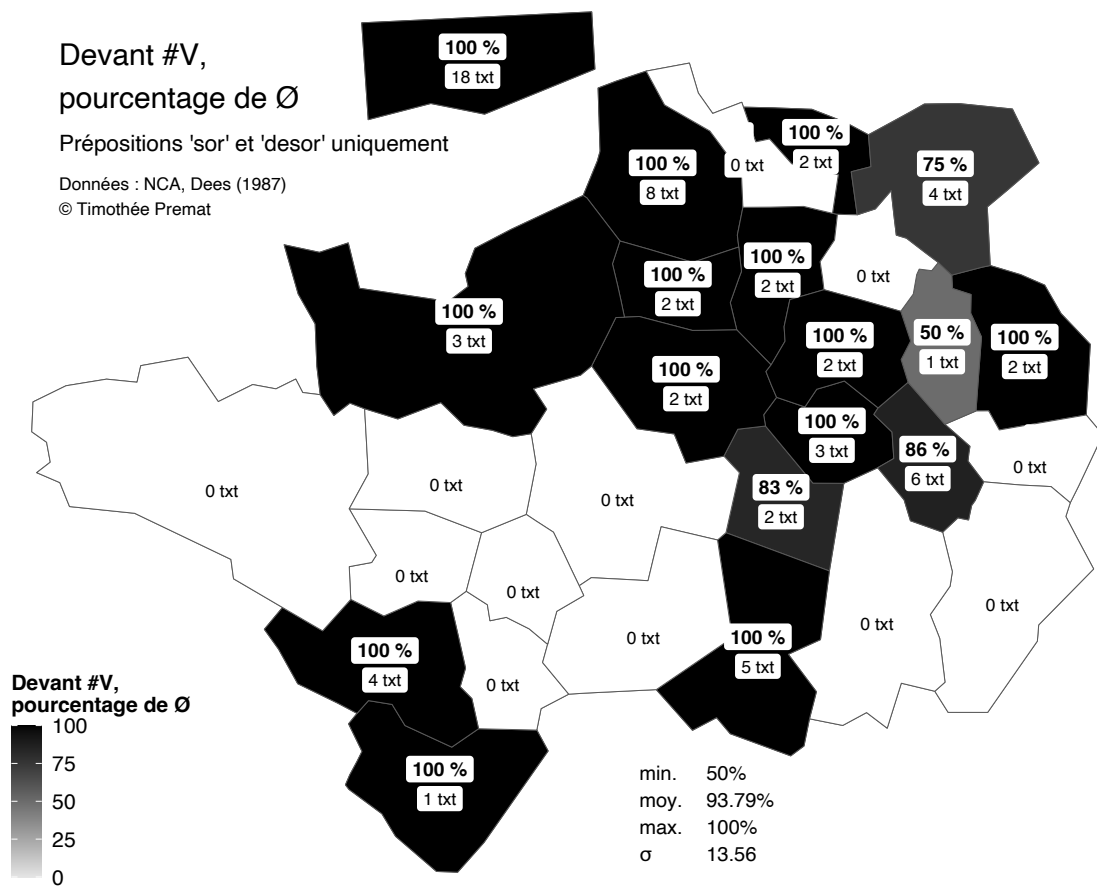


FIGURE 6.26 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, prépositions « sor » et « desor » suivies d'une initiale vocalique.

Devant initiale vocalique (fig. 6.26), la situation est bien moins homogène ($\sigma = 14$). Elle l'est pour deux raisons : d'une part de nombreuses régions deviennent muettes, le nombre d'occurrences de ces prépositions devant #V étant bien moins élevé que devant initiale consonantique (158 occ. __#V vs 3440 occ. __#C). D'autre part, alors que le taux régional moyen demeure assez proche de celui du contexte préconsonantique (94% __#V vs 96% __#C), on voit émerger des régions disposant de taux s'écartant fortement de cette moyenne, avec un taux de défaut de <e> de 50% dans la Meuse (mais sur un seul texte), et un taux de 75% en Wallonie (sur quatre textes). De manière assez significative, les autres taux régionaux, tous supérieurs à 83%, se répartissent ainsi : toutes les régions où nos lemmes sont attestés disposent de taux de 100% de défaut de <e> (avec entre 1 texte en Charente-Maritime et 18 textes en Angleterre), à l'exception de l'Yonne (83%, 2 textes) et de la Haute-Marne (86%, 6 textes). Aussi, les seules régions dont le taux est inférieur à 100% forment un arc de cercle de la Wallonie à l'Yonne, et trois régions sur quatre sont continues (Wallonie, Meuse, Haute-Marne), tandis que l'Yonne est séparée de cet ensemble, mais avec une absence d'occurrences en Bourgogne.

6.2.5.4 En somme

La distribution diatopique des défauts de <e> de nos lemmes est fortement contrastée, et ce dans plusieurs dimensions. La situation est plus homogène pour « de » que pour « entre » (devant initiale vocalique), avec un décrochage marqué de l'Angleterre uniquement pour « de », et de l'Angleterre et du Sud-Est pour « entre ». Enfin, pour « sor » et « desor », la distribution diatopique devant initiale vocalique est un peu plus contrastée que la distribution devant initiale consonantique, avec de nombreuses régions muettes et quatre régions à taux inférieurs à moyenne, qui se détachent clairement de l'ensemble des autres régions, qui ne connaissent pas de défaut de <e>. Si l'on note une certaine ressemblance entre la distribution diatopique des taux pour « de » et « entre » (détachement de l'Angleterre, faiblesse plus (« entre ») ou moins (« de ») marquée au Sud-Est), la diatopie des défauts de <e> de « contre » ne semble pas répondre aux mêmes principes, et inverse même une partie du comportement du Sud-Est. Aussi, si l'on peut postuler un principe diatopique au moins partiellement commun pour « de » et « entre », les autres lemmes semblent répondre à leur propre logique géographique.

6.2.6 Interface des distributions diatopiques et diachroniques

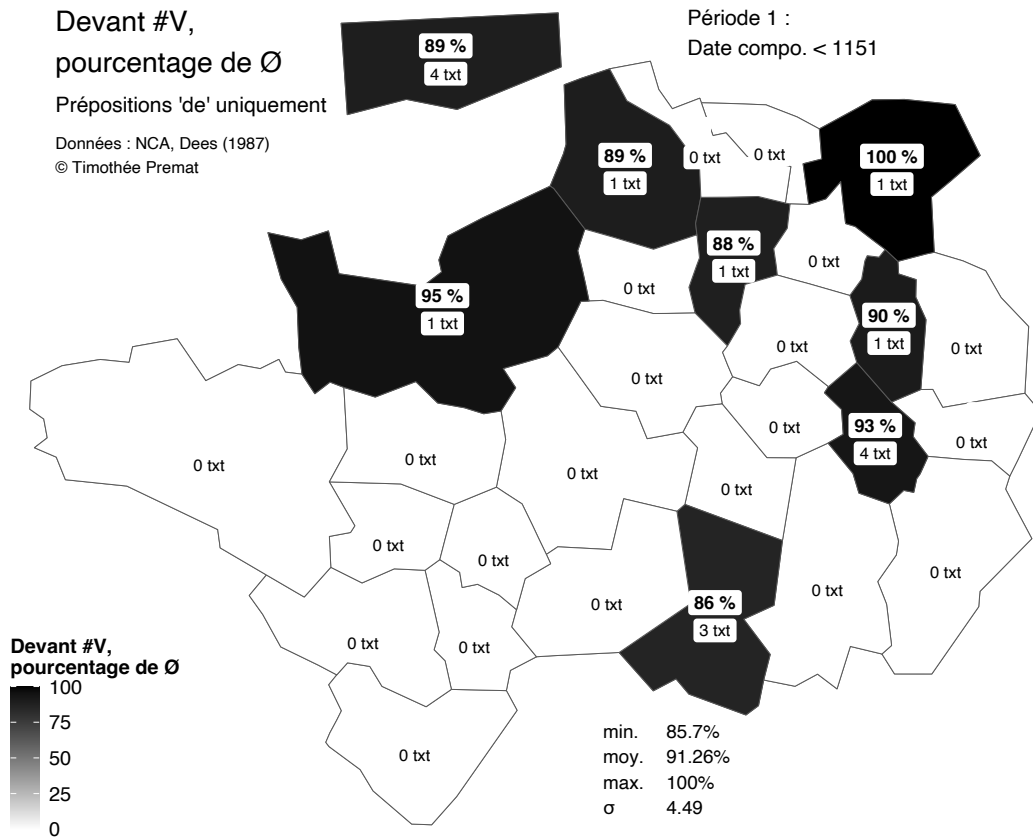
Étant donnée la variété des distributions diatopiques selon les lemmes observées *supra*, nous ne présentons pas, ici, l'interface entre diachronie et diatopie pour l'ensemble des lemmes, mais uniquement pour « de » et pour « entre ». Nous omettons de nouveau « contre » et « rencontre », dont les occurrences défectueuses sont trop rares pour soutenir une analyse diatopique, et laissons également de côté « sor » et « desor » : non seulement ces mots ne témoignent pas d'une élision (pas d'influence du contexte prévocalique), mais l'analyse diatopique synchronique a déjà montré que la diatopie de « sor » et « desor » ne rejoignait pas celle de « de » et « entre ». S'il aurait pu être intéressant de comparer l'interface entre diachronie et diatopie si « sor » et « desor » avaient partagé des points communs diatopiques pertinents avec « de » et « entre », l'analyse de cette interface est inutile en l'absence de ces ressemblances diatopiques.

6.2.6.1 « De »

Les cartes en fig. 6.27 et 6.28 en pages suivantes représentent les taux de défaut de <e> devant initiale vocalique du lemme « de », selon les périodes utilisées pour les parties du discours précédentes, c'est-à-dire des demi-siècles pour les trois premières périodes, et près de 75 ans pour la dernière période. Étant donnée la forte homogénéité du domaine continental pour ce lemme, nous ne traitons pas chaque période dans une section dédiée, mais traitons d'abord la situation en ancien français continental puis en anglo-normand.

Domaine continental. Sur toutes les périodes, le domaine continental est homogène et très largement dominé par les défauts de <e>. Dès la seconde moitié du XII^e s., on peut observer l'apparition d'une bande

a



b

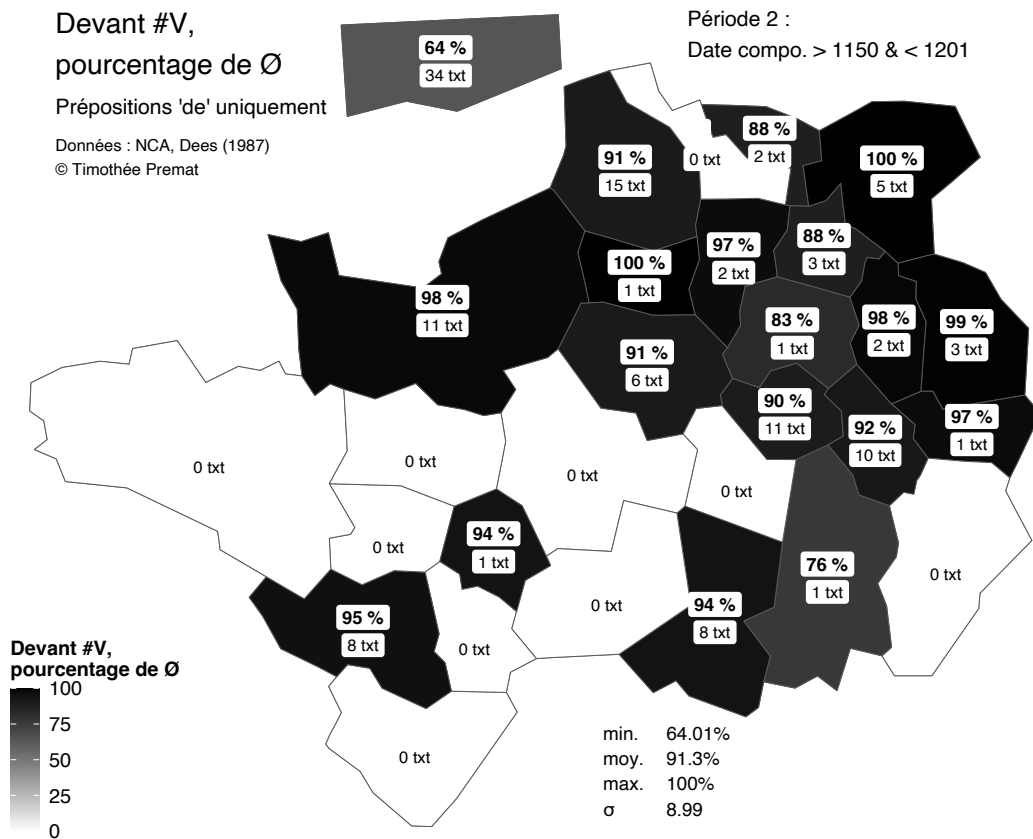
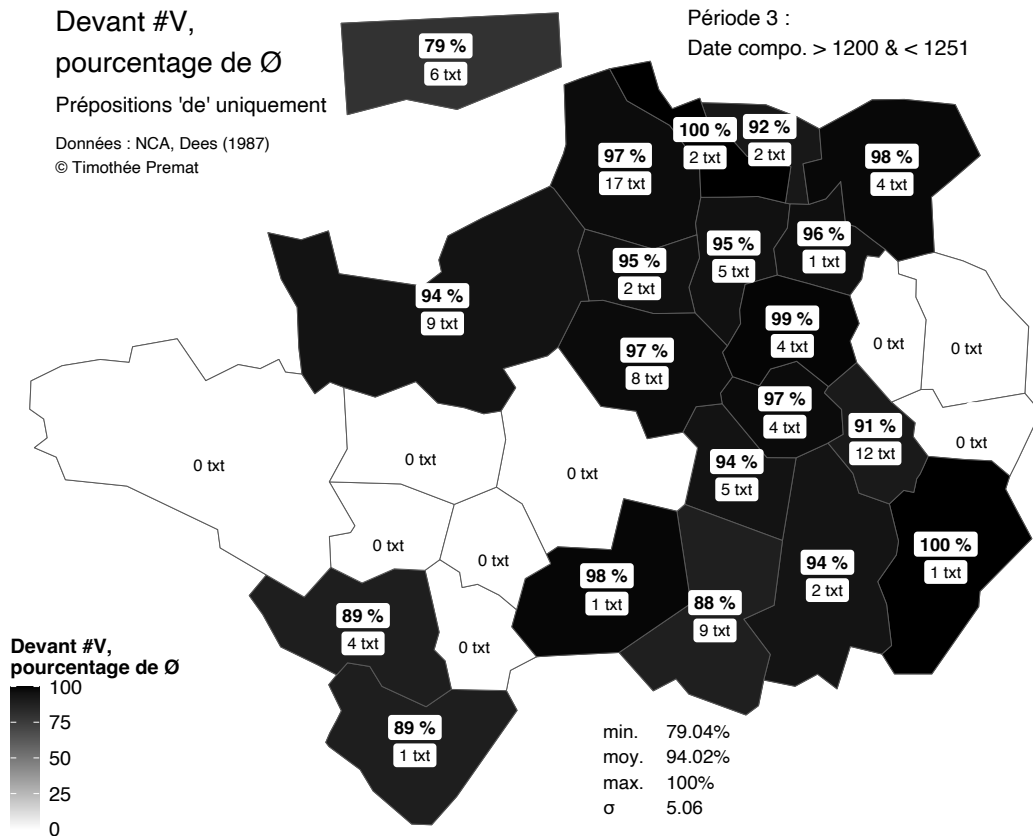


FIGURE 6.27 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1150) et 2 (b : 1151-1200), prépositions « de ».

a



b

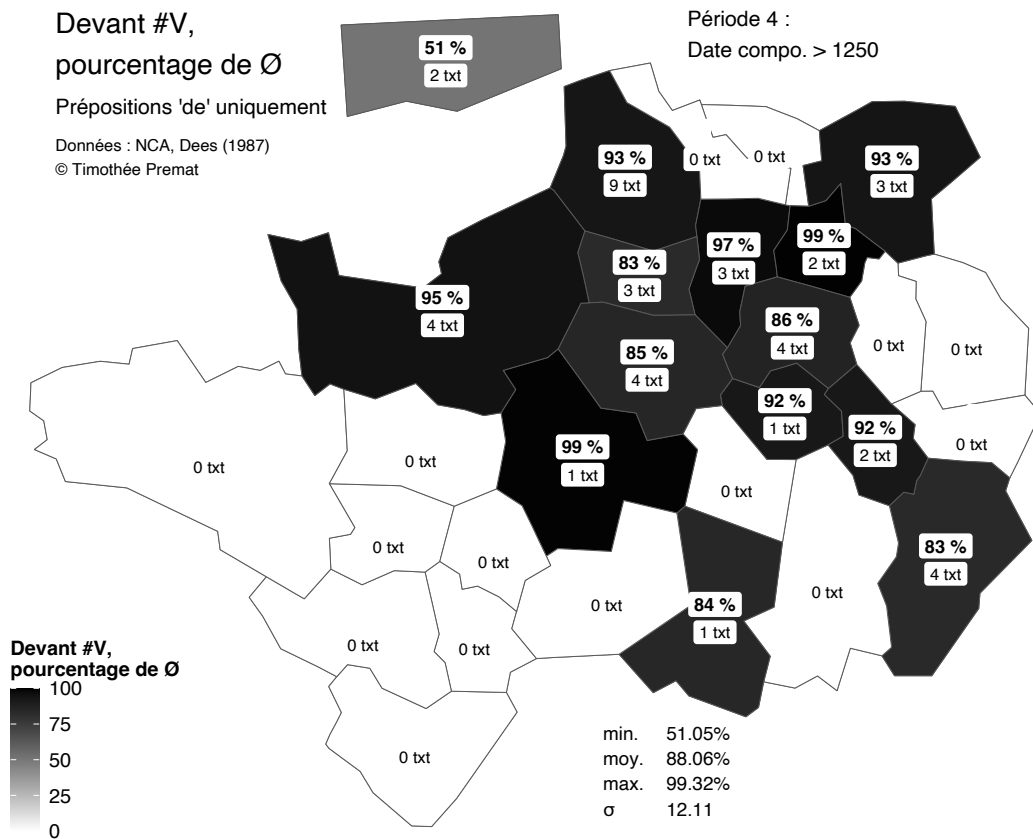


FIGURE 6.28 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), prépositions « de ».

partant de la région parisienne en direction du Sud-Est où les taux de défaut de <e> sont plus faibles. Néanmoins, cette bande disparaît dans la première moitié du XIII^e s., où seule une partie de la marge méridionale dispose de taux inférieurs à 90% (Charente-Maritime et région Vendée, Deux-Sèvres : 89%, région Nièvre, Allier : 88%).

Dans la seconde moitié du XIII^e s., les taux les plus faibles ré-apparaissent au centre du domaine (Oise : 83%, région parisienne : 85%, Marne : 86%) tandis que la Bourgogne confirme son faible taux (84%) et que la Franche-Comté accède elle aussi à un taux faible (84%, sur 4 textes), alors que son taux était de 100% dans la première moitié du siècle (mais portant sur un seul texte). Aussi, la répartition que nous avons notée sur la carte synchronique *supra* ne vaut qu'en synchronie et n'apparaît, en diachronie, que par fragments ; c'est la combinaison de ces fragments diachroniques qui compose la distribution synchronique, avec des effets d'alternance selon lesquels des régions peuvent passer d'un taux maximum à un taux inférieur à la moyenne d'une période à l'autre, et *vice-versa*. La situation est donc fort instable, ne saurait représenter une dynamique claire, et évoque plutôt une dimension stochastique dans la variation à l'œuvre.

Angleterre. L'Angleterre est également sujette à une certaine variabilité selon les périodes. Ainsi, dans la première moitié du XII^e s., elle ne dispose pas d'un taux anormalement bas (89%) vis-à-vis de la moyenne des régions (91%), ni du taux le plus bas du domaine (région Nièvre, Allier : 86%). Ce n'est qu'à partir de la seconde moitié du XII^e s. que l'Angleterre acquiert les taux de défaut de <e> les plus bas du domaine, avec un taux de 64% alors que la moyenne des régions est de 91%. Dans la première moitié du XIII^e s., son taux est toujours significativement plus bas (79%) que les taux continentaux (minimum sur le continent : 88%) et bien inférieur à la moyenne des régions (94%), mais l'écart est plus réduit que dans la seconde moitié du XII^e s.

Enfin, dans la seconde moitié du XIII^e et au début du XIV^e s., l'écart entre l'Angleterre et le domaine continental se creuse de nouveau, et de manière extrême : l'Angleterre connaît des défauts de <e> dans la moitié de ses occurrences, alors que la valeur minimale sur le continent est de 83% (Franche-Comté et Oise) et que la moyenne de toutes les régions est de 88%. Cette évolution des taux de défaut de <e> en Angleterre selon les périodes suit une dynamique exactement inverse à celle que nous avons notée pour les adjectifs féminins singuliers.

TABLE 6.3 – Taux de défaut de <e> (Tx.Ø) devant initiale vocalique en anglo-normand, adjectifs féminins singulier et prépositions « de ».

Période	Tx.Ø ADJFÉM Sg	Nb. textes	Tx.Ø PRÉP « de »	Nb. textes
(1) 1100-1150	8%	4	89%	4
(2) 1151-1200	21%	28	64%	43
(3) 1201-1250	5%	6	79%	6
(4) 1251-1322	37%	2	51%	2

La table 7.7 permet une comparaison de l'évolution des taux de ces deux catégories pour chaque période, et montre l'inversion de la dynamique entre les catégories : lorsque le taux de défaut des adjectifs augmente, il

diminue pour « de », et lorsqu'il diminue pour les adjectifs, il augmente pour les prépositions.

6.2.6.2 « entre »

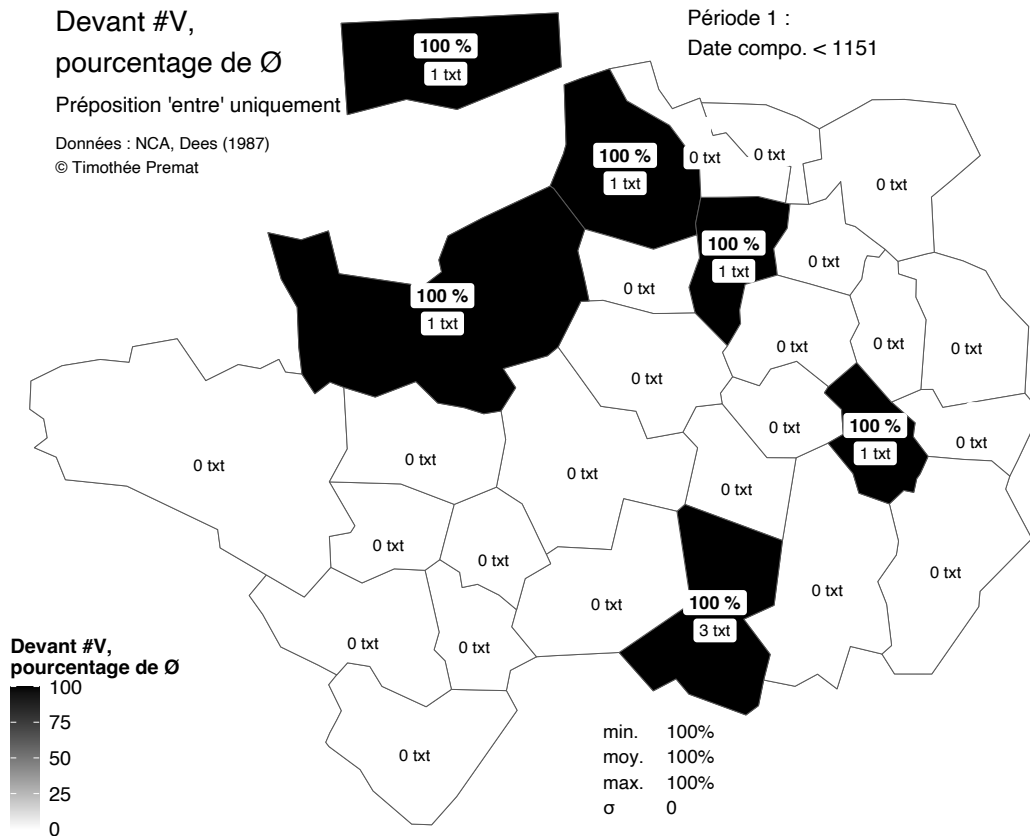
Les cartes en fig. 6.29 et 6.30 en pages suivantes représentent la distribution diatopique des taux de défaut de <e> pour le lemme « entre », pour chacune des 4 périodes ici étudiées. Comme indiqué en fig. 6.29.a, la forme défectueuse *entr* est la seule à apparaître devant initiale vocalique dans la première moitié du ^{xiii} s. Dans la seconde moitié du siècle (fig. 6.29.b), la situation est contrastée, avec une région, la Bourgogne, ne disposant que d'occurrences sans défaut de <e> (mais sur un seul texte), et quatre régions disposant de taux inférieurs à la moyenne régionale, l'Angleterre (47%, 26 textes), les Ardennes (50%, 1 texte), la Bourgogne (60%, 5 textes) et l'Aisne (69%, 1 texte). Aussi, le décrochage de l'Angleterre est visible dès cette période, et est consolidé par le nombre de textes, alors que les autres taux inférieurs à la moyenne sont moins consolidés par le nombre de textes ; ils ne transparaissent pas sur la carte synchronique en fig. 6.24.

Dans la première moitié du ^{xiii} s. (fig. 6.30.a), l'Angleterre décroche de manière encore plus visible (4%), mais certes avec un nombre de textes plus réduit (4 textes). Elle est rejointe par deux régions continentales qui disposent d'une ou plusieurs occurrences sans défaut de <e> mais ne connaissent pas d'occurrences en défaut de <e>, la Wallonie et l'Oise ; mais ces régions n'ont qu'un seul texte comportant ce lemme dans ce contexte, aussi leur taux peut n'être qu'un artefact. Le reste du domaine est très contrasté, avec des taux allant de 50% (Aube, 2 textes) et 53% (Haute-Marne, 6 textes) à 100% (Charente-Maritime, région Indre, Cher, Ardennes). Ces régions n'ayant que des formes en défaut de <e> ne disposent que d'un seul texte, ce qui rend le caractère absolu de leurs défauts de <e> suspect ; on trouve en revanche des régions disposant de taux très élevés et disposant d'un nombre raisonnable de textes, comme la Normandie (95%, 7 textes) et la région parisienne (92%, 7 textes). Aussi, si l'obtention d'un taux absolu de défaut de <e> (100%) ou d'absence de défaut de <e> (0%) semble être associée strictement à la présence d'un seul texte dans la région, la présence de taux très élevés (Normandie, région parisienne) ou très bas (Angleterre) n'est pas nécessairement liée à un faible nombre de textes. La première moitié du ^{xiii} s. forme donc un ensemble peu homogène sur le plan diatopique ($\sigma = 34$).

Enfin, dans la seconde partie du ^{xiii} s. et au début du ^{xiv} s. (fig. 6.30.b), l'Angleterre ne connaît plus de défaut de <e>, et rejoint en cela la Wallonie ; dans les deux cas, le nombre de textes attestant notre lemme en contexte prévoicalique est extrêmement faible (2 textes pour l'Angleterre, 1 pour la Wallonie). Le reste du domaine est peu homogène, mais avec une certaine concentration des taux supérieurs à la moyenne sur un court axe Sud-Ouest - Nord-Est. Le manque d'homogénéité de l'ensemble du territoire est équivalent à celui du demi-siècle précédent ($\sigma = 33$).

Vis-à-vis de la carte synchronique en fig. 6.24, le décrochage de l'Angleterre et du Sud-Est ne se met en place qu'à partir de la seconde moitié du ^{xii} s. et se réduit au ^{xiii} s., tandis que la Wallonie, la région Somme, Pas-de-Calais et l'Oise ne décrochent qu'à partir de la première moitié

a



b

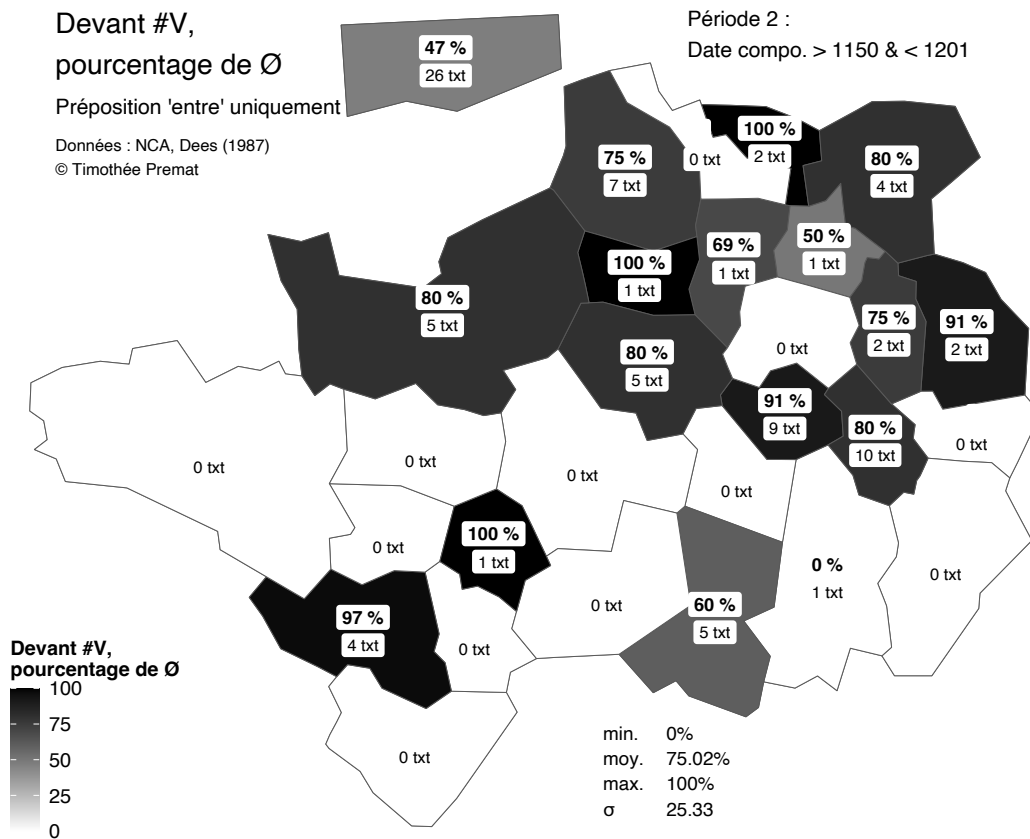
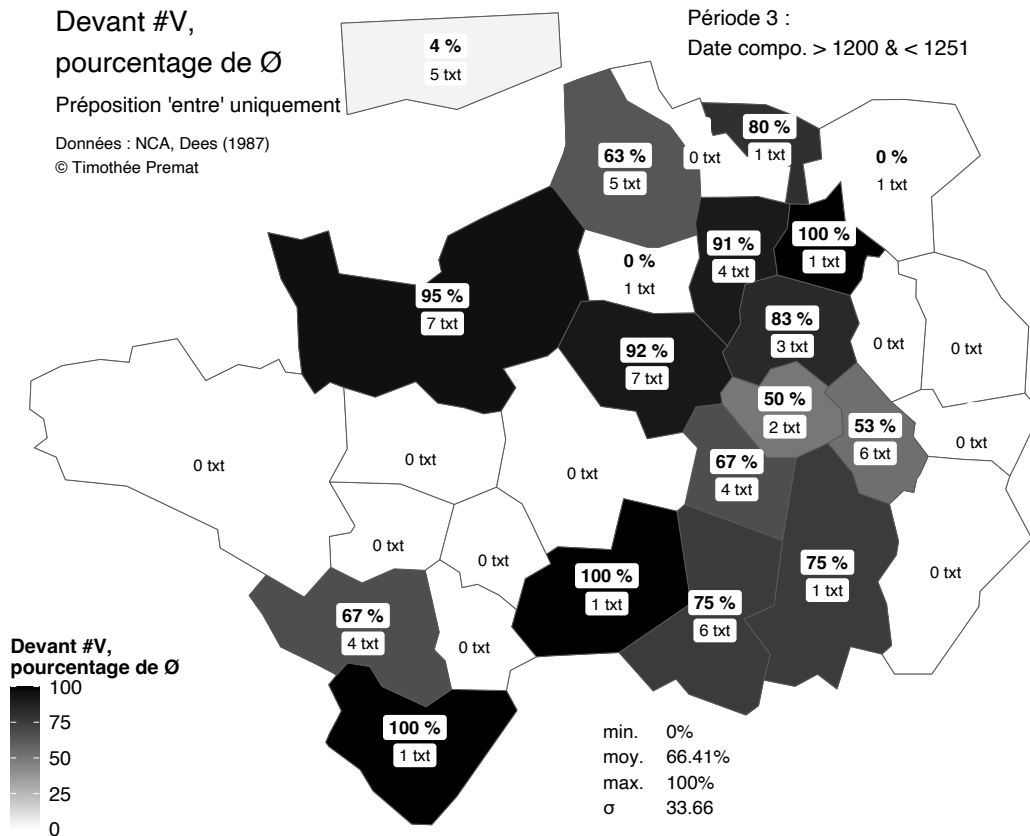


FIGURE 6.29 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1150) et 2 (b : 1151-1200), prépositions « entre ».

a



b

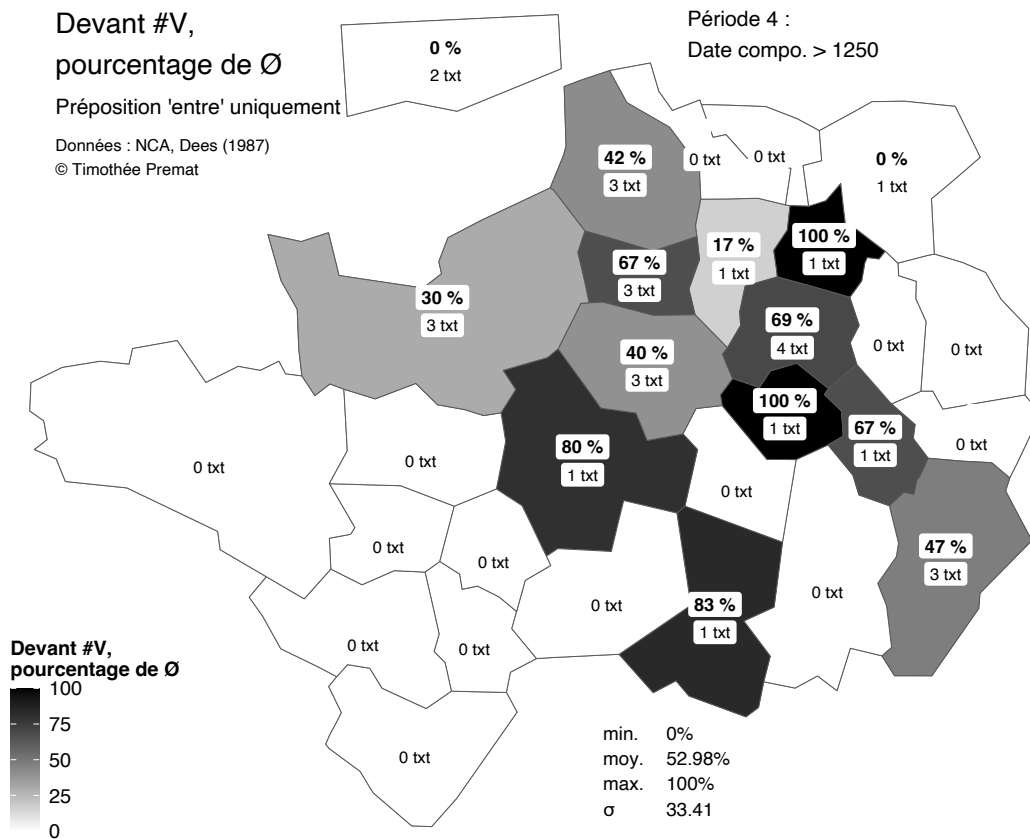


FIGURE 6.30 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), prépositions « entre ».

du ^{xiii} s. Il s'ensuit que, pour le Sud-Est, la représentation synchronique n'est, une nouvelle fois, que la fusion de situations diachroniques contrastées ; en revanche, le décrochage de l'Angleterre vaut à partir de la seconde moitié du ^{xii} s., tandis que la faiblesse de la Wallonie, de la région Nord, Pas-de-Calais et de l'Oise sont valables dès le début du ^{xiii} s. Il y a donc quelques continuités diachroniques au sein de cet ensemble plutôt hétérogène. Enfin, l'Angleterre connaît une décroissance continue (100% → 47% → 4% → 0%), qui s'oppose clairement aux évolutions en dents de scie qu'elle connaît pour les adjectifs féminins singuliers et pour la préposition « de ».

6.2.6.3 En somme

Pour la préposition « de », l'Angleterre ne se détache du domaine continental que dans la seconde moitié du ^{xii} s. et l'évolution de ses taux demeure en dents de scie. Par ailleurs, si certaines zones du centre et du Sud-Est du domaine témoignent de taux inférieurs à la moyenne, ces taux ne sont pas constants de période en période, et sont difficilement analysables. Pour « entre », le décrochage de l'Angleterre est cette fois-ci continu, tandis que les taux continentaux sont hétérogènes, avec la formation, au ^{xiii} s. d'une bande Sud-Ouest - Nord-Est qui concentre la plupart des taux élevés. Néanmoins, même au sein du ^{xiii} s., il y a peu de continuité régionale de période en période.

6.3 Des observations aux généralisations

Dans cette section, nous proposons de résumer la profusion de données exposée *supra* à des principes généraux et qui soit sont analytiques soit peuvent servir de support à une analyse. Nous commençons par synthétiser ce que nous avons pu relever au niveau du comportement distinct des lemmes et des effets de collocation (sec. 6.3.1), avant de revenir sur l'effet du contexte sur la génération de formes en défaut de <e> (sec. 6.3.2), puis sur l'évolution diachronique (sec. 6.3.3) et sur la distribution diatopique (sec. 6.3.4) des données. L'intersection de ces deux derniers points nous amène à étudier, comme pour les adjectifs féminins singuliers, le statut spécifique de l'anglo-normand et son influence sur les dynamiques de l'ensemble du jeu de données (sec. 6.3.5). Enfin, nous terminons en replaçant nos six lemmes dans l'ensemble des prépositions, pour envisager l'étendue de la variation au sein de la catégorie (sec. 6.3.6).

6.3.1 Lemmes atteints par le défaut de <e>

Nous avons retenu six lemmes atteints par le défaut de <e>. Nos six lemmes peuvent être rangés en deux catégories : ceux qui se terminent (ou devraient se terminer) par C_ə (« de », « sor », « desor ») et ceux qui se terminent par TR_ə, où TR représente une attaque branchante (« entre », « contre », « encontre »). Dans le premier cas, aucun autre lemme prépositionnel qui ne soit pas une erreur d'annotation ne se termine par C_ə ; il est donc naturel que seuls « de » et « sor » et « desor »

fassent l'objet d'une alternance <e>/Ø. Dans le second cas, un autre lemme se termine par un groupe TRə, « outre », mais il est rarement prépositionnel,²⁷ mais d'autres lemmes se terminent pas des groupes biconsonantiques + schwa autres que TR (p.ex. /st/ pour « joste » et ses préfixés) et Comparalem n'a pas détecté de variation pour ces autres prépositions. La spécificité de « entre », « contre » et « rencontre » semble donc tenir non pas à leur seule phonologie, mais à un paramètre dans l'emploi de ces prépositions. Ces deux éléments peuvent être résumés sous la forme des généralisations en (6.79) et (6.80).

(6.79) **PRÉP : sélectivité lexicale.**

Les seuls lemmes en Cə et TRə susceptibles de connaître le défaut de schwa le connaissent, mais des lemmes en C.Cə pourraient le connaître et ne le connaissent pas.

(6.80) **PRÉP : sélectivité syntaxique.**

D'autres lemmes en CCə (mais pas en OLə) pourraient connaître le défaut de <e>, mais seuls « entre », « contre » et « rencontre » le connaissent. Cela peut provenir de configuration syntaxique et/ou fréquentielles (cf. 6.83) propres à ces lemmes.

Diachroniquement, nous n'avons pas noté de diffusion lexicale progressive qui ne soit pas le fait de l'évolution de la taille du corpus. Les occurrences *d*, *entr* et *sur* sont présentes dès *Roland*, tandis qu'« rencontr », « contr » et « desor » apparaissent au milieu du XII^e s. ou dans la seconde moitié de ce siècle, ce qui correspond simplement au moment où la taille du corpus devient suffisamment importante pour permettre l'attestation de variables minoritaires. Cet élément est repris en (6.81).

(6.81) **PRÉP : diffusion lexicale.**

Il n'y a pas de phénomène de diffusion lexicale du défaut de <e> observable.

Par ailleurs, le trop faible nombre de lemmes retenus empêche toute évaluation du rapport entre fréquence d'emploi et taux de défaut de <e>, d'autant que ce dernier dépend surtout du contexte pour trois lemmes sur cinq (6.82).

(6.82) **PRÉP : fréquence.**

Il est inapproprié de chercher un effet de la fréquence d'emploi sur les taux de défaut de <e> de cinq lemmes.

Enfin, si le statut syntaxique des collocations implique nécessairement un effet fréquentiel de *type* fort (PRÉP + syntagme prépositionnel), il ne semble pas actif pour « de », « sor » et « desor », dans la mesure où le défaut de <e> est la règle, pour ces prépositions, devant initiale vocalique pour la première, et dans tous les contextes pour les deux autres. En revanche, les prépositions « entre », « contre » et « rencontre » semblent réagir assez fortement à leur association avec le pronom personnel masculin de troisième personne au cas régime pluriel (« eux »), ce qui va de pair, pour « contre », avec la transformation de la préposition en préfixe dans de nombreuses dérivations verbales, nominales et adjectivales. Les taux de défaut associés à ces constructions PRÉP + « eux » varient fortement : on compte 395 occurrences du type « entr eus » pour 125 occurrences

27: Deux des trois occurrences *outr* et *otr* annotées comme prépositionnelles sont des locutions (*otr mesure* et *otr mer, roma*, p. 12, 21).

du type *entre eus*, six occurrences du type *encontr eus* pour 1 occurrence du type *encontre eus*, tandis que l'on ne trouve que 6 occurrences du type *entr eus* pour 26 occurrences du type *entre eus*. Par ailleurs, les effets dus à la grammaticalisation où à la seule fréquence de *type* ne se limitent pas strictement à la combinaison avec le PROPER3MASCpl CR, et d'autres constructions (*entre* + verbe, « contre » + « aval » ou « amont ») sont attestées en défaut de <e>, leur statut de *type* voire de *token* étant attesté soit à bas bruit dans les données, soit dans les dictionnaires.

La prédominance très marquée de la construction avec le pronom « eux » peut s'expliquer par la dimension sémantique de la préposition. Tant « entre » que « contre » prédisposent le pronom à apparaître dans un *type* syntaxique plus restreint que le type syntaxique prépositionnel par défaut 'PRÉP + syntagme prépositionnel', à savoir « Prép + ProPerCR » et, dans le cas de « entre », « Prép + ProPerCR »pl. À cela s'ajoute le fait que « eux » est le seul pronom à initiale vocalique qui apparaît régulièrement derrière ces prépositions. De ce fait, le type syntaxique semble pouvoir induire un fonctionnement en *token*, c'est-à-dire en unité indivisible, ce qui expliquerait par deux dimensions les défauts de <e> : d'une part le REDUCING EFFECT qui tend à supprimer des segments au sein d'un *token* (BYBEE 2006, p. 714-715, 724-725, 2000, 2002 ; BYBEE et SCHEIBMAN 1999), et d'autre part la fusion des constituants (BYBEE 2006, p. 715), qui abolit la frontière entre les deux mots et crée, ici, un hiatus əV.

Néanmoins, contrairement à ce que nous avons pu développer en sec. 5.3.1 pour les adjectifs féminins singuliers, il ne s'agit probablement pas ici d'un processus de grammaticalisation (*entr els* ou *contr els* ne témoignent pas de la même grammaticalisation que *contredire* ou *contrefait*). Il s'agit alors simplement d'une modalité cognitive du traitement des unités linguistiques, qui, selon certaines théories — ce point est controversé (DIVJAK 2019, p. 143) — permet un traitement différencié des unités grammaticalisées et des constructions récurrentes mais non grammaticalisées. Dans ce cas, c'est bien l'association de deux constituants qui est activée (*contr els*), au lieu que soit activée directement l'association d'un constituant indivisible (*contredire*, *contrefait*). Même si les constructions fréquentes non grammaticalisées (telles que *contr els*) n'ont pas le statut des constructions grammaticalisées, c'est-à-dire qu'elles ne sont pas sujettes au même traitement cognitif, elles n'ont pas non plus le statut d'une association syntaxique peu courante, étant notamment caractérisées par une vitesse de traitement supérieure ; ce qui peut évoquer, pour notre analyse phonologique, un seul domaine phonologique plutôt que plusieurs. Au demeurant, la fréquence des défauts de <e> dans les constructions « entre » + verbe ne peut évoquer qu'un fonctionnement de *type* et non de *token*, le deuxième élément de la construction étant libre tant au niveau de son lemme que de sa flexion. Pour les occurrences *contr aval* et *contr amont*, en revanche, il y a lieu de postuler une grammaticalisation classique, ce que font d'ailleurs de nombreux dictionnaires qui présentent ces formes non comme des locutions mais comme de simples lemmes. Le constat de l'association relativement récurrente entre défauts de <e> de « entre », « contre » et « rencontre » et le pronom « eux », avec des verbes et, de manière plus rare mais confortée par les sources lexicographiques, avec les prépositions²⁸ « aval » et « amont » est résumé, sans ses implications théoriques, en (6.83).

BYBEE (2006). « From Usage to Grammar: The Mind's Response to Repetition »
 — (2000). « The phonology of the lexicon: Evidence from lexical diffusion »
 — (2002). « Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change »
 BYBEE et SCHEIBMAN (1999). « The effect of usage on degrees of constituency: the Reduction of *don't* in English »

DIVJAK (2019). *Frequency in Language: Memory, Attention and Learning*.

28: Dans les emplois en question, « aval » et « amont » sont des prépositions (GD); le fr.mod. *en aval* et *en amont* serait un pléonisme en a.fr.

(6.83) **PRÉP : collocations.**

Le défaut de <e> de « entre », « contre » et « encontre » semble être favorisé par la collocation avec le PROPERS3pl CR « eux » et, à plus faible fréquence, avec d'autres éléments.

6.3.2 Le rôle du contexte

L'effet du contexte dépend fortement du lemme : quatre lemmes ne connaissent de défauts que devant initiale vocalique (« de », « entre », « contre » et « encontre »), tandis que deux lemmes ne semblent pas sensibles au contexte (« sor » et « desor »). Le contexte prévocalique a un effet extrêmement marqué sur la génération de défauts de <e> du lemme « de » ($t = 86$), encore bien marqué pour le lemme « entre » ($t = 23$), tandis que cet effet est bien moins marqué pour « contre » et « encontre » ($t = 5$); dans tous les cas, la mesure est hautement significative ($p < 0,005$). Pour ces lemmes, cet effet signifie qu'on ne trouve pas, ou en quantité vraiment marginale, d'occurrences défectueuses devant initiale vocalique. Au niveau de l'interprétation linguistique, la puissance de l'effet (ici mesurée par la valeur- t) indique, dans le cas de « de », que le défaut de <e> est la règle (taux moyen de défaut par texte $_ \#V$: 95%), ce qui est aussi le cas mais de manière moins marquée pour « entre » (taux moyen de défaut par texte $_ \#V$: 68%), tandis que pour « contre » et « encontre », les formes défectueuses demeurent marginales (taux moyen par défaut texte $_ \#V$: 24%). Enfin, si la condition nécessaire aux défauts de <e> représentée par le contexte prévocalique est stable sur l'ensemble des régions et des périodes, la puissance de génération de défauts de <e> de ce contexte varie dans le temps et dans l'espace, et particulièrement en anglo-normand. Pour « de », ces variations dans le temps sont principalement dues à une spécificité de l'anglo-normand, qui est sur-représenté dans la seconde moitié du XII^e s., tandis que pour « contre » et « encontre », ces variations tiennent principalement à la faible quantité de données disponibles (126 occurrences). Cette faible quantité de données pour ces deux lemmes crée à la fois des vides dans les données ainsi que des conditions qui empêchent l'émergence de formes défectueuses (puisque minoritaires) et, parallèlement, renforcent leur effet lorsque celles-ci apparaissent (puisque non compensées par de nombreuses formes non défectueuses). Ces éléments sont résumés en (6.84) et (6.85).

(6.84) **PRÉP : contexte, « de », « entre ».**

Le contexte prévocalique est le seul à être associé à des défauts de <e>, et cette association est majoritaire.

(« De » : $t = 86$; $p < 0,005$; moy. $_ \#V = 89\%$)

(« Entre » : $t = 23$; $p < 0,005$; moy. $_ \#V = 68\%$)

L'effet n'est pas stable dans le temps et l'espace.

(6.85) **PRÉP : contexte, « (en)contre ».**

Le contexte prévocalique est le seul à être associé à des défauts de <e>, et cette association est minoritaire.

($t = 4,62$; $p < 0,005$; moy. $_ \#V = 24\%$)

La stabilité de l'effet ne peut être mesurée étant donnée la faible quantité des occurrences.

Pour « sor » et « desor », en revanche, la valeur-*t* comparant les deux contextes est proche de zéro ($t = 0,02$) et la mesure est non significative ($p = 0,98$), ce qui signifie que l'hypothèse nulle (absence d'influence du contexte) ne peut être rejetée. Analytiquement, cette absence d'influence du contexte n'est pas seulement possible, elle est probable, et ce d'autant plus que « sor » et « desor » se démarquent clairement des autres lemmes étudiés. Ici, la moyenne des taux de défaut de <e> est sensiblement la même dans les deux contextes (95%) et, hormis quelques variations locales à des périodes données qui semblent dues au hasard, cet absence d'effet est stable dans l'espace et le temps. Linguistiquement parlant, cette absence d'effet du contexte couplée à une moyenne très élevée traduit une probable ré-analyse de la forme sous-jacente, qui du *soure* /surə/ archaïque et prédictible depuis l'étymon latin, est devenue /sur/ ou /sor/ en perdant son schwa final. Ces éléments sont résumés en (6.86).

(6.86) **PRÉP : contexte, « (de)sor »**

Le contexte phonologique n'a pas d'effet sur les défauts de <e> des PRÉP « sor » et « desor », dont le <e> semble avoir majoritairement disparu.
($t = 0,02$; $p = 0,98$).

Certaines grammaires (dont FOUCHÉ 1969 [1958], p. 513; POPE 1966 [1934], p. 118, 218) proposent de lire dans cette ré-analyse une généralisation de l'élosion (/surə/ → [sur] / __#V généralisé en /surə/ → /sur/). Nos données ne permettent pas de confirmer cette hypothèse, dans la mesure où le taux moyen non pondéré de défaut de <e> est sensiblement le même dans les deux contextes : 95%; mais la lexicalisation de la forme sans /-ə/ a pu précéder notre corpus. Nos données ne permettent pas non plus de réfuter l'hypothèse, mais les mesures qui pondèrent les taux par le nombre d'occurrences par texte peuvent induire un certain doute : le taux moyen pondéré de défaut de <e> est très légèrement supérieur devant initiale consonantique (97%) au taux moyen pondéré devant initiale vocalique (96%), et l'homogénéité des données est meilleure devant initiale consonantique ($\sigma = 0,09$) que devant initiale vocalique ($\sigma = 0,15$). Surtout, il faut relever que les lemmes « sor » et « desor » apparaissent 22 fois plus souvent devant initiale consonantique (3340 occ.) que devant initiale vocalique (158 occ.), ce qui rend difficile d'imaginer une réforme de la représentation sous-jacente du lemme à partir de sa forme de surface dans un contexte marginal. É. BOURCIEZ et J. BOURCIEZ (1974 (1967), p. 89) proposent plutôt une analogie avec *sus* (← *sūsu* ← *sūrsu*); POPE (1966 [1934], p. 216) n'évoque cette analogie que comme facilitant la rephonologisation. CLÉDAT (1917, p. 11-12) s'en tient à noter que la finale atone peut chuter « avant l'accent » (italiques de la source).

Enfin, il faut rappeler que notre analyse n'a porté que sur les occurrences annotées pos="PRE" de ces deux lemmes, et que nous avons pu constater que cette annotation était parfois imparfaite, dans les deux dimensions : certaines occurrences sont annotées pos="PRE" alors qu'elles ne devraient pas l'être, et d'autres ne sont pas annotées pos="PRE" alors qu'elles devraient être considérées comme des prépositions. De ce fait, ces résultats sont à prendre avec précaution.

FOUCHÉ (1956). *Phonétique historique du français — Les voyelles.*

POPE (1966 [1934]). *From Latin to Modern French with Especial Consideration of Anglo-Norman.*

É. BOURCIEZ et al. (1974 (1967)). *Phonétique française, Étude historique.*

CLÉDAT (1917). *Manuel de phonétique et de morphologie historique du français.*

6.3.3 Diachronie générale

Si, sur l'ensemble des lemmes et des contextes, le taux de défaut de <e> est relativement stable, avec seulement une très légère décroissance, cette stabilité masque des situations très hétérogènes selon le contexte et les lemmes. Si l'on s'en tient à observer chaque contexte, le taux de <e> décroît dans le temps devant initiale consonantique, alors qu'il croît devant initiale vocalique. Dans la mesure où seules les prépositions « sor » et « desor » connaissent des défauts de <e> devant initiale consonantique, ce sont elles seules qui peuvent être responsables de la décroissance du taux de défaut de <e> en contexte préconsonantique ; or leur propre taux de défaut de <e> ne décroît pas de manière marquée dans le temps. C'est donc le rapport quantitatif entre « sor » et « desor » et les autres lemmes qui, devant initiale préconsonantique, est responsable de cette décroissance du taux de défaut de <e> en contexte préconsonantique. Devant initiale vocalique, la croissance du taux de défaut de <e> devant initiale vocalique est due à la préposition « de », principalement par élimination progressive des valeurs les plus basses, c'est-à-dire des textes dans lesquels le défaut de <e> n'est pas généralisé. Ce point est repris par la généralisation (6.87).

(6.87) **PRÉP : diachronie (a).**

Par élimination progressive des *outliers*, la préposition « de » fait augmenter le taux de défaut de <e> devant initiale vocalique. Il n'y a pas probablement pas d'évolution diachronique pertinente dans le contexte préconsonantique.

Enfin, devant initiale vocalique et, dans une moindre mesure devant initiale consonantique, on observe que les taux de défaut de <e> culminent temporairement. Cet effet est peu visible devant initiale consonantique, la courbe en légère décroissance prenant légèrement la forme d'une courbe en 'S', avec un premier plateau au cours du XII^e s. qui décroche à la charnière avec le XIII^e s., avant que les taux ne se stabilisent de nouveau au cours du siècle. Cet effet semble être dû, au-delà de la question du poids relatif des prépositions « sor » et « desor » vis-à-vis des trois autres lemmes, à la raréfaction des *outliers*, avec notamment une certaine concentration de textes dont le taux de défaut de <e> pour ces deux lemmes est inférieur à 100% autour de la date normalisée de 1200. Devant initiale vocalique, le 'pic' de concentration des taux les plus élevés est décalé dans le temps, tant vis-à-vis du premier plateau du contexte préconsonantique, que vis-à-vis du pic dans la seconde moitié du XII^e s. que nous avons observé pour les adjectifs féminins singuliers ; les taux de défaut de <e> culminent, ici, dans la première moitié du XIII^e s. avec, sur le graphique de série temporelle sans pondération des taux par le nombre d'occurrences des textes, une forme de plateau assez marquée sur l'ensemble du demi-siècle (cf. fig. 6.12 et agrandissement en fig. 6.13). Ces éléments sont synthétisés en (6.88), qui renvoie à (6.93 *infra*) pour la raison probable de cet effet.

(6.88) **PRÉP : diachronie (b).**

Devant initiale vocalique, on observe un plateau ou un pic, selon que les taux ne sont pas ou sont pondérés par la quantité d'occurrences, dans la première moitié du XIII^e s. (cf. 6.93).

6.3.4 Diatopie

Pour l'ensemble des lemmes retenus et en synchronie, la diatopie semble assez unifiée. Devant initiale consonantique, le domaine est fort homogène ($\sigma = 2$), avec une moyenne des taux moyens régionaux basse (4%), uniquement due aux prépositions « sor » et « desor ». Devant initiale vocalique, c'est le domaine continental qui est fort homogène, avec des taux légèrement moins élevés mais pas aberrants sur un axe Centre - Sud-Est; en revanche, l'Angleterre se détache clairement avec un taux de défaut de <e> de 67% alors que le taux continental minimal est de 85%.

Étant donné la différence quantitative entre « de » (11 291 occ.) et les autres lemmes (« entre » : 639 occ. ; « contre » et « encontre » : 126 occ.) devant initiale vocalique, ce sont les défauts de <e> devant initiale vocalique du lemme « de » qui sont en grande partie responsables de la cartographie générale. Le décrochage de l'Angleterre se confirme pour tous les lemmes, mais avec « de » il est plus marqué du fait de l'homogénéité très forte du domaine continental (fig. 6.23). Ce traitement spécifique de « de » __#V en Angleterre est d'autant plus marqué que les prépositions « sor » et « desor », dont le défaut de <e> se généralise peu importe le contexte, n'ont jamais de <e> prévocative en Angleterre (fig. 6.26).

Pour « de », au sein du domaine continental, l'axe Centre - Sud-Est dont les taux sont légèrement plus faibles apparaît dans la seconde moitié du XII^e s., disparaît dans la première moitié du XIII^e s., avant de réapparaître à la fin du corpus. Discontinu, cet axe n'est guère visible que dans la seconde moitié du XII^e s. Cette continuité géographique est donc discontinue dans le temps; elle est dotée d'une certaine cohérence qui n'induit pas à penser que seul le hasard la motive, mais la diachronie la rend discontinue, ce qui pose problème. Le domaine anglo-normand, en revanche, se détache de l'ensemble continental dès la seconde moitié du XII^e s. Nous avons noté *supra* (cf. table 7.7) que son évolution, toujours en dents de scie, était exactement inverse à celle observée pour les adjectifs féminins singuliers : son taux croît entre la seconde moitié du XII^e s. et la première moitié du XIII^e s., avant de décroître à la fin de la période. Pour « entre », en revanche, la décroissance du taux de défaut de <e> en anglo-normand est constante, montrant qu'il ne s'agit pas d'une propriété uniforme de l'a.n. L'ensemble de ces éléments est synthétisé en (6.89).

(6.89) **Prép : diatopie.**

Devant initiale consonantique, le domaine est très homogène, et les manquements temporaires à cette homogénéité semblent n'être dus qu'au hasard.

Devant initiale vocalique, le domaine continental est assez homogène, avec simplement un axe Centre - Sud-Est dont les taux sont légèrement moins forts dans la seconde partie des XII^e et XIII^e s. L'Angleterre se détache clairement du domaine continental à partir de la seconde moitié du XII^e s.

6.3.5 Poids de l'anglo-normand et de son hétérogénéité

Si l'anglo-normand ne connaît pas de spécificité devant initiale consonantique, sa spécificité est clairement visible devant initiale vocalique. Pour

nos quatre prépositions sensibles au contexte, et comme représenté en fig. 6.31, l'anglo-normand est responsable non seulement de la présence d'une quantité relativement élevée de textes dont les taux de défaut de <e> sont anormalement bas, mais également de tous les textes, sauf un, dont le taux de défaut de <e> est inférieur à 50%. Comme déjà évoqué à propos des adjectifs féminins singuliers, les textes anglo-normands se concentrent dans la seconde moitié du XII^e s. et, ce qui est intéressant ici, c'est que cette concentration se fait avec des valeurs plus faibles : les textes anglo-normands entre 1150 et 1200 forment une 'barre verticale' qui s'étend sur toute la hauteur de l'axe des ordonnées, et dont le centre de gravité se situe hors de l'étendue de la dispersion des données continentales à la même époque.

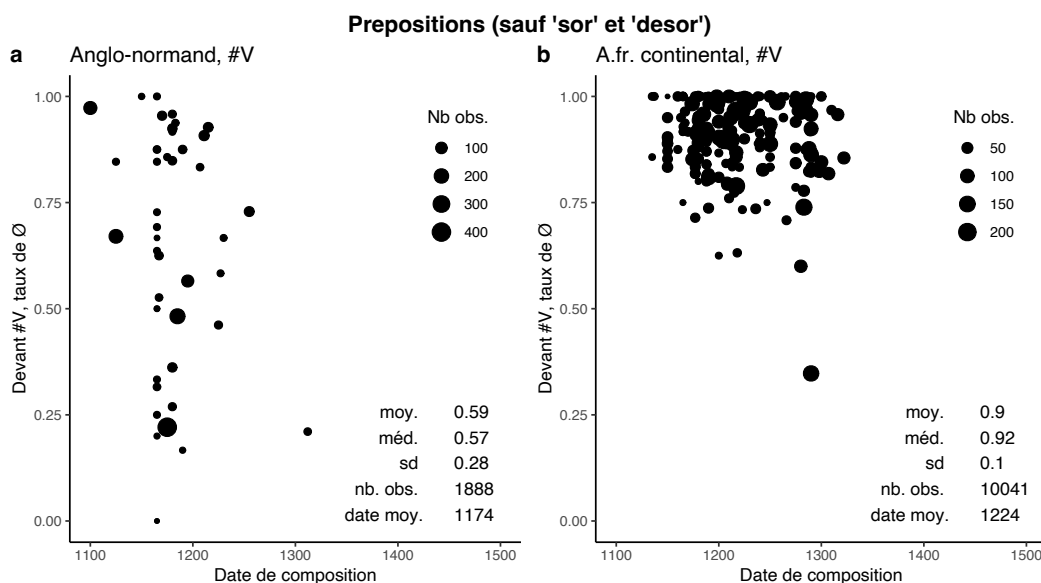


FIGURE 6.31 – Diachronie des taux de défaut de <e>, l'anglo-normand vs ancien français continental, prépositions « de ».

Cette spécificité de l'anglo-normand est si forte qu'elle imprime sa marque sur l'ensemble de la période. Ainsi, le taux moyen de défaut de <e> en anglo-normand est bien inférieur au taux moyen continental (6.90). Cette situation est analogue, bien qu'inverse, à celle que nous avons notée pour les adjectifs féminins singuliers. Par ailleurs, et encore une fois comme pour les adjectifs féminins singuliers, la spécificité de l'anglo-normand ne tient pas qu'à son taux moyen de défaut de <e>, mais également à l'éclatement des données, dont la dispersion est bien plus importante que sur le domaine continental (6.91).

(6.90) PRÉP²⁹ : a.n. moy. = 59% vs a.fr.cont. moy. = 90%
 ADJFÉM_{Sg} : a.n. moy. = 11% vs a.fr.cont. moy. = 3%³⁰

(6.91) PRÉP³¹ : a.n. σ = 0,28 vs a.fr.cont. σ = 0,1
 ADJFÉM_{Sg} : a.n. σ = 0,16 vs a.fr.cont. σ = 0,07³²

29: Sauf « sor » et « desor ».

30: Il s'agit des taux pondérés par le nombre d'occurrences par texte, comme en fig. 6.31.

31: *Idem* n. 29.

32: *Idem* n. 30.

Enfin, il faut noter que cette spécificité de l'anglo-normand ne vaut, en réalité, que pour « de ». Comme décrit auparavant, le taux anglo-normand de défaut de <e> pour le lemme « entre » est en décroissance constante, tandis que le domaine continental est bien plus contrasté qu'il ne l'est pour « de ». Ainsi que le montre la comparaison des graphiques de gauche en fig. 6.32, si l'on retrouve bien cet effet de 'barre verticale' de

0 à 100% pour « entre » dans la seconde moitié du XII^e s., le siècle suivant est caractérisé par un seul texte disposant d'occurrences défectueuses alors que de nombreux textes ne disposent plus que d'occurrences dont le <e> est noté. Par comparaison, pour « de », les données anglo-normandes mélangent des taux élevés et des taux moyens au cours du XIII^e s. L'effet diachronique que l'on peut attribuer à la présence de textes anglo-normands nombreux dans le corpus dans la seconde moitié du XII^e s. est donc limité au lemme « de », dont les occurrences sont de toute manière bien plus nombreuses que celles des autres lemmes.

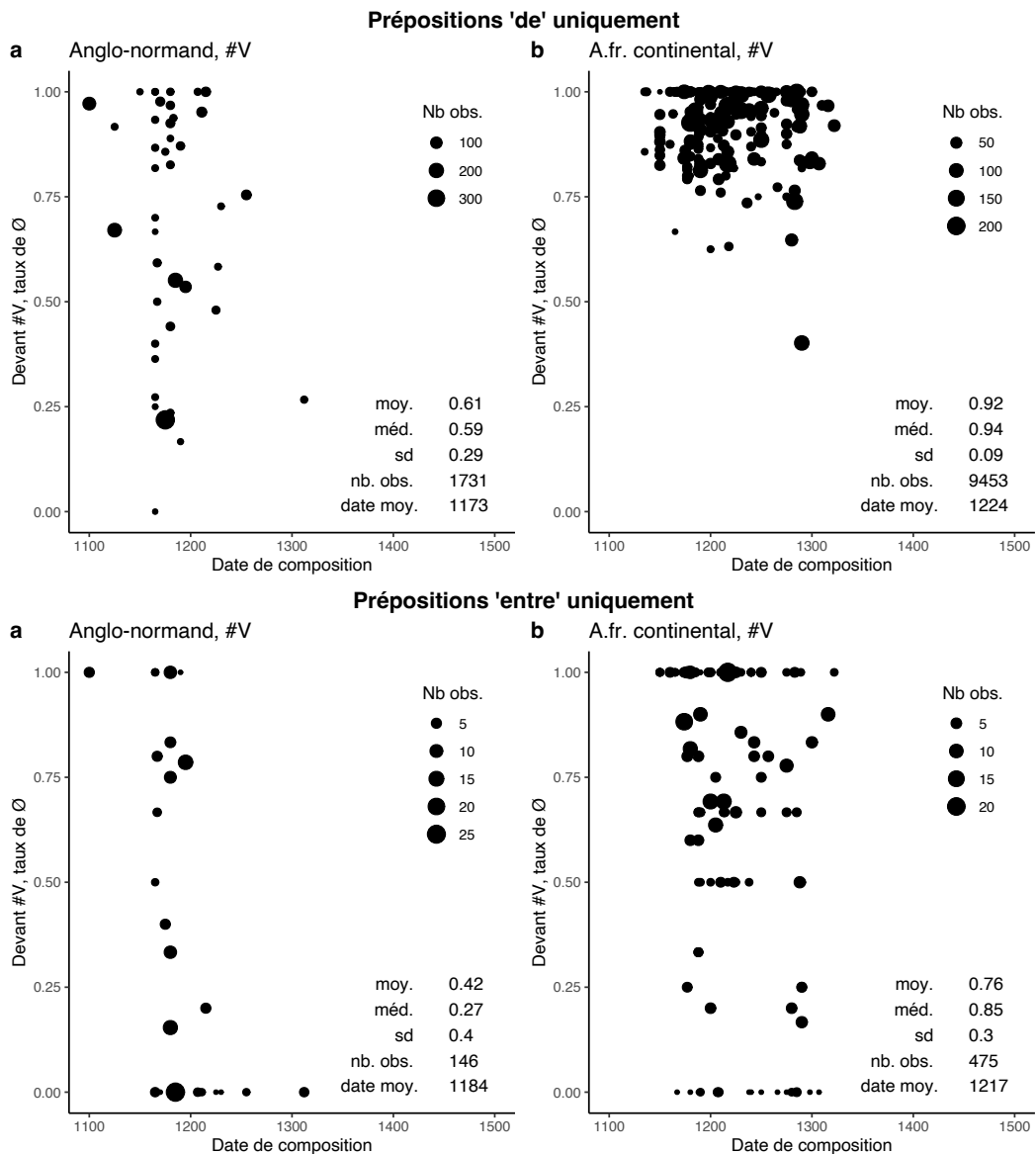


FIGURE 6.32 – Diachronie des taux de défaut de <e>, l'anglo-normand vs ancien français continental, prépositions « de » et « entre » devant initiale vocalique.

Cette hétérogénéité de l'anglo-normand a donc pour résultat d'aboutir à des taux de défaut de <e> sur « de » devant initiale vocalique plus faibles en anglo-normand que sur le continent. Cet effet, étant donné la forte dispersion des données, n'est pas un effet particulièrement fort ($t = -5,33$), mais il est statistiquement significatif ($p < 0,005$). Comme représenté en fig. 6.33, il correspond au fait que le deuxième quartile de l'anglo-normand s'étend là où le français continental ne dispose que de quelques valeurs extrêmes, tandis que le premier quartile de l'anglo-normand correspond à des valeurs tout simplement absentes sur le continent (sauf un point).

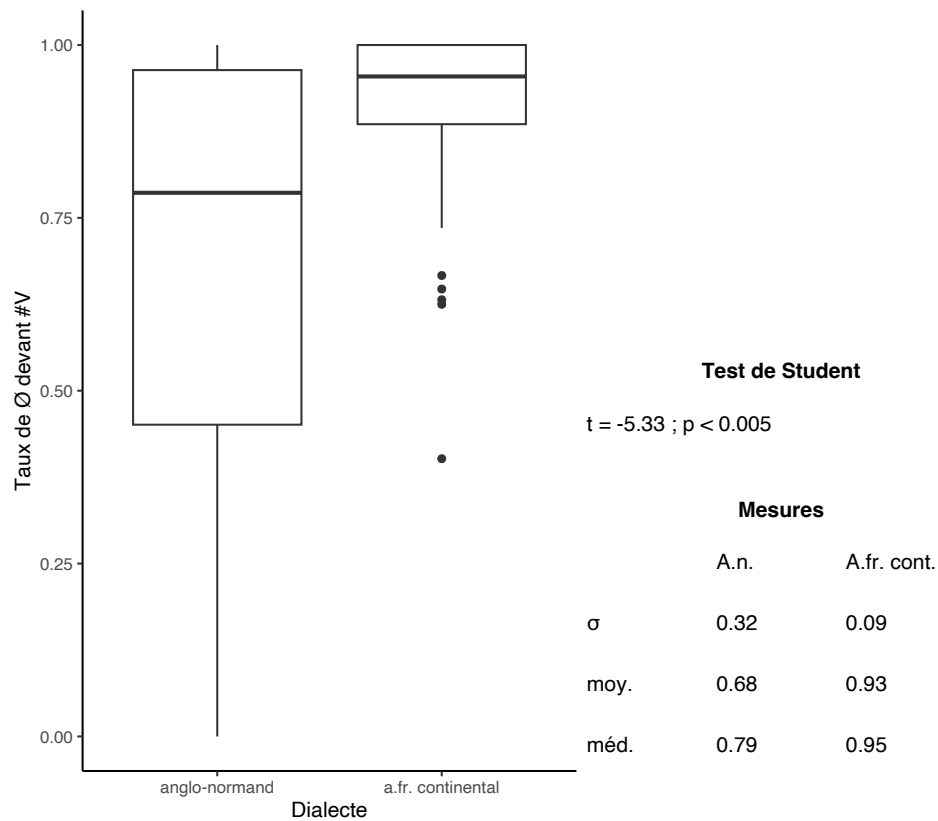


FIGURE 6.33 – Distinction entre anglo-normand et dialectes continentaux pour les taux de défaut de <e> des prépositions « de » devant initiale vocalique. Ici, les taux affichés et pris en compte pour le test de Student ne sont pas pondérés par le nombre d'occurrences des lemmes dans chaque texte.

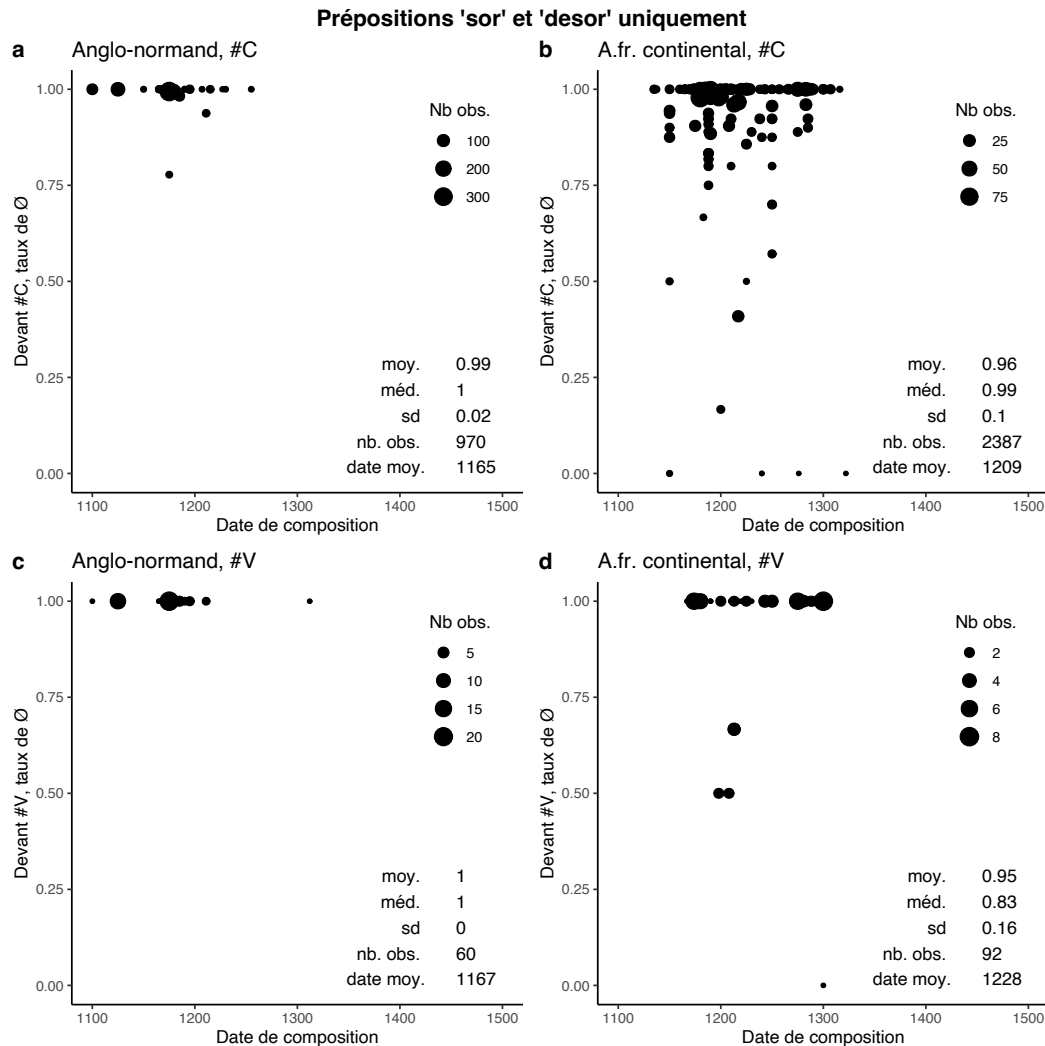


FIGURE 6.34 – Diachronie des taux de défaut de <e>, l'anglo-normand *vs* ancien français continental, prépositions « sor » et « desor ».

En revanche, pour les lemmes « sor » et « desor », comme représenté en fig. 6.34, l'anglo-normand montre une homogénéité bien plus élevée que celle de l'ancien français continental devant initiale consonantique (a.n. $\sigma = 0,02$; a.fr.cont. $\sigma = 0,1$), et ne connaît aucun cas d'absence de défaut de <e> devant initiale vocalique. L'ensemble de ces éléments est synthétisé en (6.92).

(6.92) **PRÉP : spécificité de l'a.n.**

« De » : l'a.n. (a) se distingue des dialectes continentaux par (b) son hétérogénéité et (c) ses nombreux textes à faibles taux de défaut de <e>.

- a) $_ \#V : t = -5.33 ; p < 0,005 ;$
- b) $_ \#V : \sigma = 0,32$ *vs* a.fr. continental $\sigma = 0,09 ;$
- c) $_ \#V : \text{moy.} = 68\%$ *vs* a.fr. continental $95\%.$

« (De)sor » : l'a.n. (a) est plus homogène que l'a.fr.cont., et (b) pratique systématiquement le défaut de <e> devant initiale vocalique.

- a) $_ \#V : \sigma = 0$ *vs* a.fr. continental $\sigma = 0,16 ;$
 $_ \#C : \sigma = 0,02$ *vs* a.fr. continental $\sigma = 0,1.$
- b) $_ \#V : \text{moy.} = 100\%$ *vs* a.fr. continental $\text{moy.} = 96\%.$

Ces éléments permettent de comprendre le plateau observé au XIII^e s. devant initiale vocalique, et reproduit une nouvelle fois en figure marginale 6.35. Ce plateau représente une situation dans laquelle l'anglo-normand est marginal et, dans la première moitié du siècle, doté d'un taux qui se rapproche du taux continental. De ce fait, ce plateau représente la situation normale sur le continent, et ce sont les taux plus bas avant le plateau qui représentent une situation diatopique radicalement différente, dominée par l'anglo-normand. La montée vers le plateau ne représenterait donc pas une évolution diachronique, mais représenterait uniquement un changement dans l'équilibre diatopique du corpus. Ce paramètre est analogue à celui que nous avons relevé pour les adjectifs féminins singuliers, mais son effet est inverse : alors que l'anglo-normand était responsable du pic de défaut de <e> des adjectifs féminins singuliers dans la seconde moitié du XII^e s., il est cette fois-ci responsable du taux relativement bas de défaut de <e> de la préposition « de » à cette même période. Cet élément est résumé en (6.93).

(6.93) **PRÉP : poids de l'a.n. dans les données.**

Étant donné (6.92 « de »), plus il y a de textes anglo-normands, moins les défauts de <e> devant initiale vocalique sont nombreux.

6.3.6 Place de la variation au sein des PRÉP

Toutes les mesures citées *supra* concernent les défauts de <e> des occurrences des six lemmes concernés par un défaut de <e>. Comme indiqué en table 6.1 *supra*, ces six lemmes représentent 36% des occurrences prépositionnelles du corpus. Ainsi, les taux de formes défectueuses que nous avons présentés ne sont pas aussi marginaux, au sein de l'ensemble de la catégorie morphosyntaxique, qu'ils le sont pour les noms ou les adjectifs féminins singuliers. Le taux global de défaut, sur nos six lemmes, est 16%, ce qui correspond à 6% de toutes les prépositions. De plus, ces six lemmes contiennent trois lemmes se terminant par une seule consonne et un schwa, et ce sont les seuls, parmi tous les lemmes prépositionnels, à se terminer en C₀ : tous les lemmes prépositionnels en -C₀ sont donc sujets à des défauts de <e>. Les trois autres lemmes se terminent par un groupe *obstruante + liquide (+ schwa)* ; le seul autre lemme de préposition en OL₀, *outre*, n'est que très rarement en emploi prépositionnel et est le plus souvent adverbial. Si l'on ne peut exclure des emplois prépositionnels de « outre » non annotés comme tel (et que nous ne pouvons donc vérifier), on peut néanmoins considérer que tous les lemmes prépositionnels se terminant en C₀ et tous, ou presque tous, les lemmes se terminant en OL₀ sont atteints par le défaut de <e>. Ces éléments sont repris en (6.94)

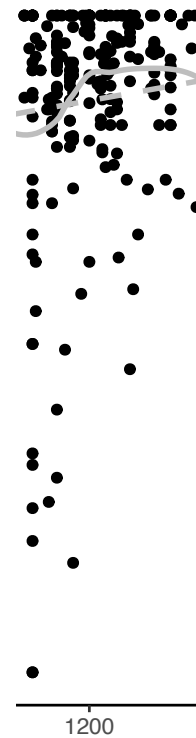


FIGURE 6.35 – Extrait de la fig. 6.12.b, montrant le plateau de la régression locale autour de 90% entre 1200 et 1250.

(6.94) **PRÉP : poids de la variation.**

Rapportés à l'ensemble des PRÉP, les défauts de <e> sont globalement minoritaires (a), mais concernent tous les lemmes prépositionnels se terminant en Cə et OLə (b). Leur statut varie fortement en fonction des lemmes et des contextes (c).

- a) Défauts : 6% des occ. de PRÉP, 16% des occ. des 6 lemmes.
 b) Les 6 lemmes retenus contiennent 100% des lemmes en -Cə et ~100% de ceux en -OLə
 c) « de » 0% __#C 89% __#V
 « entre » 0% __#C 68% __#V
 « (en)contre » 0% __#C 24% __#V
 « (de)sor » 95% __#C 95% __#V

6.4 Conclusion

Dans ce chapitre, nous avons étudié la distribution des défauts de <e> des prépositions dans le corpus du NCA. Après avoir, dans l'introduction, présenté les exclusions de données (sec. 6.1.1) et les formes relemmatisées (sec. 6.1.2), nous avons présenté les résultats de l'analyse que Comparalem nous a permis de mener (sec. 6.2), et avons proposé une brève typologie préliminaire des lemmes et formes sujets à défauts de (sec. 6.2.1). Après avoir brièvement abordé les principales caractéristiques de la distribution de ces défauts de <e> en diachronie (sec. 6.2.2), nous avons développé la spécificité de chaque lemme par ordre d'apparition dans le corpus, et avons traité des quelques cas de collocations (sec. 6.2.3), avant de traiter de la distribution des défauts de <e> selon le contexte phonologique (sec. 6.2.4), en diatopie (sec. 6.2.5), et à l'interface entre diachronie et diatopie (sec. 6.2.6). Enfin, nous avons procédé à une montée en généralisation, qui tient lieu de *discussion* des résultats observés (sec. 6.3).

Cette généralisation nous a permis de réduire les faits observés à un ensemble de caractéristiques majeures. Au niveau des lemmes (sec. 6.3.1), nous avons noté que les lemmes concernés étaient les seuls lemmes prépositionnels à se terminer en Cə (6.79) mais que d'autres lemmes se terminant en TRə ou CCə ne connaissaient pas de défaut de <e>. (6.80). Nous n'avons pas noté de phénomène de diffusion lexicale observable dans nos données qui ne soit pas dû à la simple taille du corpus (6.81), ni de corrélation entre la fréquence des lemmes et leurs taux de défaut de <e> (6.82). En revanche, pour les lemmes « entre », « contre » et « encontre », nous avons montré que les structures PRÉP + « eux » semblaient particulièrement productives de défauts de <e>, ce pronom étant par ailleurs le seul pronom personnel à initiale vocalique à apparaître régulièrement après ces prépositions (6.83).

L'analyse contextuelle a révélé des propriétés différentes en fonction des lemmes (sec. 6.3.2). Le défaut de <e> du lemme « de », presque systématique hors du domaine anglo-normand devant initiale vocalique, est absent devant initiale consonantique (6.84). De même, les lemmes « entre », « contre » et « encontre » ne connaissent pas de défaut de <e> devant initiale consonantique, mais leurs défauts sont moins majoritaires (« entre ») voire sont minoritaires (« contre », « encontre ») devant initiale vocalique (6.85). Enfin, les lemmes « sor » et « desor » montrent un défaut de <e> généralisé, et relativement insensible au contexte (6.86).

L'analyse diatopique (sec. 6.3.4) a montré que la force de l'effet contextuel était variable pour les prépositions « de » et « entre » : si l'Angleterre ne connaît jamais de défauts de <e> en quantité significative devant initiale consonantique, cette région se signale devant initiale vocalique par un taux de défaut de <e> bien inférieur aux taux rencontrés sur le continent (6.89 ; 6.92). Diachroniquement, cet effet est le plus marqué dans la seconde moitié du XII^e s. et, pour « de » mais pas pour « entre », sur la fin du corpus. Sur le continent, on note simplement un axe Centre-Sud-Est où les taux sont légèrement plus faibles, mais sans commune mesure avec le décrochage de l'Angleterre à partir du milieu du XII^e s. Ce décrochage de l'Angleterre au milieu du XII^e s., associé au fort nombre de textes anglo-normands à cette période, explique le creux qui précède le plateau des taux de défaut de <e> de « de » devant initiale vocalique (6.93) ; il ne s'agit donc pas d'un effet diachronique mais d'un effet diatopique.

Enfin, nous avons noté que nos cas de défaut de <e> étaient loin d'être tout à fait marginaux au sein de l'ensemble des prépositions (sec. 6.3.6). Si les défauts de <e> ne concernent certes que 6% des occurrences prépositionnelles, ce chiffre est 10 fois plus élevé que pour les adjectifs féminins singuliers (0,57%). Par ailleurs, ces 6% représentent 16% des occurrences de nos lemmes, et ces lemmes comportent les seuls lemmes qui se terminent en C₀ et TR₀ (6.94). Singulièrement, nous n'avons pas trouvé de défaut de <e> pour les lemmes se terminant par un autre groupe consonantique suivi de schwa (p.ex. « joste », « decoste »). Aussi, nous concluons que les défauts de <e> analysés pour nos cinq lemmes prépositionnels sont significatifs. Pour « sor » et « desor », ils ne sont pas sensibles au contexte et indiquent une réanalyse phonologique générale de /sorə/ en /sor/ ; pour « de », « entre », « contre » et « encontre », ils ne peuvent survenir devant initiale consonantique et, devant initiale vocalique, ils sont presque systématiques hors du domaine anglo-normand pour « de », tandis qu'ils semblent dépendre de conditions syntaxiques et de fréquences d'association avec le mot suivant pour « entre », « contre » et « encontre ».

7.1 Introduction

Ce chapitre est dédié à l'analyse de la variation graphique <e>/Ø dans les adverbes pour lesquels <e> est de droit, étant prédictible à partir de l'étymologie et des règles de l'évolution phonétique. Nous entendons montrer que les adverbes représentent la seule catégorie pour laquelle le défaut de <e> n'est ni extrêmement marginal ni conditionné par une initiale vocalique subséquente. En effet, si le défaut de <e> demeure relativement marginal pour l'ensemble de la catégorie (6% d'occurrences adverbiales en défaut de <e>, comme pour les prépositions, cf. 6.2), il est dominant pour les 9 lemmes atteints par cette variation, avec un taux moyen de défaut par texte de 68% devant initiale vocalique et de 74% devant initiale consonantique.

Dans ce chapitre, nous commençons par proposer une note sur la pérennité de certaines alternances adverbiales <e>/Ø, étant donnée la vitalité de ces alternances après le Moyen Âge (sec. 7.1.1). Nous abordons ensuite les exclusions de lemmes et de formes (sec. 7.1.2) et les quelques fusions de lemmes que nous avons opérées (sec. 7.1.3), avant de présenter les résultats concernant les défauts de <e> (sec. 7.2) et de discuter ces résultats pour les doter d'une portée analytique et monter en généralité (sec. 7.3).

7.1.1 Note sur la pérennité des alternances <e>/Ø (et <es>) des adverbes

Les adverbes connaissent encore une importante variation <e>/Ø aux ^{xvi}^e et ^{xvii}^e s. et parfois jusqu'au ^{xviii}^e s., attestée par les pratiques métriques et par le discours des grammairiens de l'époque. Cette variation n'est pas strictement limitée aux défauts de <e> : elle concerne aussi des adverbes qui sont exclus de nos données, et notamment tous ceux dont la finale *-que* est due à l'analogie avec *ŪNQVAM* → *onque* (*adonc* ~ *adonque(s)*; *donc* ~ *donque(s)*; etc.; POPE 1966 [1934], p. 118-119; REY 2012, TLFi) ou procède de *ŪSQUE* (cf. sec. 7.1.2.1). Ces lemmes ne font pas partie de nos données puisque leur <e> n'est pas de droit, fût-il par la suite intégré à la forme sous-jacente du mot ou non (cf. sec. 7.1.2).

Que le <e> soit de droit ou non, la variation <e>/Ø d'une partie des adverbes se transmet donc au moins jusqu'en français préclassique, du moins dans la langue des vers. Cette variation est souvent liée à la présence d'un '-s adverbial' non étymologique (MARCHELLO-NIZIA 2020a, GGHE, p. 914); lorsque ce '-s adverbial' s'ajoute à une variation <e>/Ø, l'adverbe dispose de trois formes, au comportement métrique différent. Ainsi, la forme étymologiquement attendue *ore* (← *HÓRAM*) s'accompagne de la forme défectueuse *or* et de la forme non défectueuse avec '-s adverbial' *ores*; la première s'élide dans les vers devant initiale vocalique, tandis que la deuxième est toujours monosyllabique (et ne

POPE (1966 [1934]). *From Latin to Modern French with Especial Consideration of Anglo-Norman*.

REY (2012). *Dictionnaire historique de la langue française*.

MARCHELLO-NIZIA (2020). « Les adverbes ».

correspond qu'à une position métrique) et que la troisième ne s'élide pas et est donc toujours bisyllabique (et correspond, dans le vers, à deux positions métriques). Ces doublets ou triplets perdurent donc un certain temps, notamment dans le vers où leur statut métrique différent se maintient par licence poétique, comme en témoignent Palsgrave et Ronsard au XVI^e s. (THUROT 1881-1883, p. 181) :

By lycence poetycall, to kepe their ryme, and suche lyke fygyures use they in all other partes of speche, as *onc* for *oncques*, *donc* for *donques*; *avec* for *avecques*, and suche many other.

(PALSGRAVE 1530, p. 394)

Tu diras selon la contrainte de ton vers *or*, *ore*, *ores*, *adonc*, *adoncque*, *adoncques*, *avecq'*, *avecques* et mille autres que sans crainte tu trancheras et allongeras ainsi qu'il te plaira, gardant tousiours une certaine mesure consultée par ton aureille.

(RONSARD 1866 [1565], p. 333)

La régularisation de ces formes diffère, dans sa direction comme dans sa chronologie, selon les lemmes. Selon THUROT (1881-1883, p. 181), on doit à Deimier et Malherbe la disparition des formes défectueuses *or*, avec notamment une distinction grammaticale pour Deimier :

On dict [...] *ores*, *ore*, *or*, mais c'est en diuers sens [...] *Ore* signifie le temps present, comme [...] *maintenant* et à *present* [...] *Or* [...] est vn mot de conjoinction [...] Il ne faut iamais oster la derniere lettre de ce terme, *ore*, et dire par licence desbordee *or'*, avec ceste apostrophe [...] pour vouloir signifier le susdict temps present, veu qu'estant ainsi abregé, il denote vne autre chose, outre que les vers en sont extremement rudes.

(DEIMIER 1610, p. 188-190)

Or' pour *maintenant* ne se dit point.

(MALHERBE [1862], p. 308, 463)

Au contraire, pour « onque », Deimier préconise la forme bisyllabique au détriment de la forme monosyllabique, pour des raisons d'eurythmie. D'autres formes, malgré des opinions divergeantes, semblent se maintenir jusqu'au XVII^e s. Il en va ainsi du couple *avec* ~ *aveque*, dont SYLVIUS [DU Bois] (1531, p. 157) admet les deux formes à droits égaux, mais dont MEIGRET (1550b, fol. 119), RAMÉE (1572, p. 120) et ESTIENNE (1569 [1557], p. 99, 1549) ne mentionnent que la forme sans <e>, tandis que Vaugelas maintient encore les deux formes, y compris en prose (THUROT 1881-1883, p. 184) :

Avec et *avecque* sont tous deux bons, et ne sont pas seulement commodes aux poètes pour allonger ou accourcir leurs vers d'une syllabe, selon la nécessité qu'ils en ont, mais encore à ceux qui escriuent en prose avec quelque soin de satisfaire l'oreille, soit pour former la juste mesure d'une période, soit pour les joindre aux mots avec lesquels ils rendent le son plus doux, et la prononciation plus aisée, soit en fin pour empescher dans la prose la mesure des vers.

(VAUGELAS 1880 [1647], p. 424)

THUROT (1881-1883). *De la prononciation française depuis le commencement du XVI^e siècle, d'après les témoignages des grammairiens.*

PALSGRAVE (1530). *L'Esclaircissement de la langue française.*

RONSARD (1866 [1565]). « Abbrégé de l'art poétique français ».

DEIMIER (1610). *L'Académie de l'art poétique.*

MALHERBE ([1862]). *Œuvres.*

SYLVIUS [DU Bois] (1531). *In linguam gallicam Isagoge, una cum eiusdem grammatica latino-gallica, ex Hebræis, Græcis et Latinis authoribus.*

MEIGRET (1550b). *Le trętté de la gramęre française.*

RAMÉE (1572). *Grammaire.*

ESTIENNE (1569 [1557]). *Traicté de la grāmaire Française.*

— (1549). *Dictionnaire François latin, autrement dict Les mots François, avec les manieres duser diceulx, tournez en latin, corrigé et augmenté.*

VAUGELAS (1880 [1647]). *Remarques sur la langue française.*

L'académie ne supprime *avecque* de son dictionnaire qu'au xviii^e s. (éd. 1718), ayant remarqué 14 ans plus tôt qu'*avecque* n'est plus qu'une licence poétique, qu'elle déconseille par ailleurs (ACADÉMIE FRANÇAISE 1880 [1704], p. 429). La date de l'évolution, sa linéarité et/ou son degré de consensus, ainsi que le sens de son évolution (conservation ou perte du <e>) est donc différente selon les lemmes.¹ Il ne s'agit pas ici de retracer l'évolution de chaque adverbe à partir du français pré-classique, mais de noter que la variation est encore bien attestée, pour certains adverbes, dans les siècles qui suivent le Moyen Âge, et concerne autant les adverbes où <e> est de droit que ceux où <e> ne l'est pas.

7.1.2 Exclusions

Sur la base de nos requêtes contextuelles, Comparalem a identifié 81 étiquettes de lemmes présentant une variation <e>/∅. Comme pour les autres parties du discours, nous ne travaillons que sur les lemmes dont le <e> est de droit, et avons donc exclu l'ensemble des lemmes dont le <e> ne provient pas directement de l'application des règles de phonétique historique sur l'étymon. Ainsi, nous avons exclu les doublets dont la forme en <e> serait due à l'influence de ŪNQUAM (7.1 a) — en revanche, ŪNQUAM → *onque* ~ *onc* est bien pris en compte, et ceux qui présentent un <e> mais dont la source est obscure et ne saurait être étymologique (7.1 b). Nous avons également exclu les cas où l'alternance détectée par Comparalem est due à un graphème <e> final ne représentant pas un schwa (7.1 c) ou à un graphème <a> représentant la finale atone dans le Sud-Ouest (7.1 d), à une erreur d'annotation en lemmes (7.1 e) ou en parties du discours (7.1 f). Enfin, les formes qui procèdent du latin ŪSQUE et l'adverbe négatif *ne* ont également été exclus; la justification de ces exclusions fait l'objet des deux sections suivantes (sec. 7.1.2.1 et 7.1.2.2).

(7.1) Causes d'exclusions et exemples

(a) <e> sous l'influence de UNQUAM :

- « donc » : DUNC → *donc* ~ *donque*, *donques*,
- *Idem* préfixés « adonc », « idonc »,

(b) Autre source de <e> non étymologique :

- « regieres » : DE EA RE → *gieres* (FEW) :
 - La position de RE en fin de syntagme lexicalisé ne permet pas de prédire son évolution en schwa.
- « lonc » : LONGE → *lonc*, *long*, *longe*, *longes* :
 - NG est un groupe hétérosyllabique latin, il n'impose pas la rétention d'une voyelle finale (cf. LONGU → *long*, SÉGÉRAL et SCHEER 2020, GGHF, p. 341). Si la forme *longe* vient du féminin LONGA, il s'agit d'un doublet, autre raison d'exclusion.
- « jusque », « jesque », « enjusque » et « dusque » : cf. sec. 7.1.2.1.

(c) <e> pour non-/ə/ :

- « naje »² : NE EGO → *naje* ~ *najo*,
- « très » : TRANS → *trei* ~ *tre*,

ACADÉMIE FRANÇAISE (1880 [1704]). « Observations de l'académie française sur les Remarques de M. de Vaugelas ».

1: Si, dans les exemples cités, le <e> est en définitive maintenu pour *ore* (« maintenant »), où il est de droit, et déconseillé pour *avec*, dont l'étymologie (FOUCHÉ 1966, p. 650; TLFi) et l'évolution phonétique (FOUCHÉ 1966, p. 659) sont troubles, nous ne garantissons pas que ce soit le cas pour tous les adverbes et que les grammairiens aient tranché, *in fine*, en fonction de l'origine du <e>.

2: AND : « Not I! ».

- (d) <a> pour /ə/ :
- « outre1 » : *utre, otre ~ ultra, oltra* (alia, poit, calen),³
 - « mëisme » : *meisme ~ meisma* (poit),
- (e) Erreurs d'annotation en lemme :
- *ainsiqu(e)* lemmatisé « ainsi » ; *ainzqu(e)* lemmatisé « ainz » ; *deciqu(e)* lemmatisé « deci », *presqu(e)* lemmatisé « pres »,
 - *toteveieis, totefois* etc. lemmatisés « tote »,
- (f) Erreurs d'annotation en POS :
- « tresque » et « trosque » annotés en tant qu'adverbes au lieu d'être prépositionnel ou conjonctif,
 - « ne » (en2_ne), i.e. *ne* et *n*, cf. sec. 7.1.2.2.

7.1.2.1 ÚSQUE : motivation de l'exclusion

Parmi les lemmes détectés par Comparalem comme ayant une variation <e>/Ø, quatre remontent au latin ÚSQUE : « jusque », « enjusque », « jesque » et « dusque ». Selon SAMPSON (1980), les groupes consonantiques hétérosyllabiques qui retiennent une voyelle finale sont ceux qui n'existent pas en latin (p.ex. /sn/ dans ÁSĪNUM → *âne, áċĪNU* → *aisne* ; /sl/ dans PĒSSŪLUM → *pesle, pedle*, PĒNSĪLE → *poisle* ; SÉGÉRAL et SCHEER 2020, GGHF, p. 343), tandis que ceux qui sont attestés en latin entraînent la chute de la voyelle finale (p.ex. CRĪSPU → *cresp* ; HÓSTE → *ost* ; id. GGHF, p. 341), et ce critère de latinité des groupes est donné comme convainquant pour une large majorité des formes par la GGHF (p. 336, 341 Rq. 1). Les descendants de ÚSQUE ne se rattachent pas non plus au type palatalisé, et ce à double titre. D'une part, /k^w/ n'est pas sujet à la palatalisation, étant séparé de la voyelle qui déclenche la palatalisation par l'appendice labiovélaire /^w/ (GGHF *id.*, p. 241). D'autre part, la palatalisation aboutit elle aussi à la chute de la finale atone : /jC/ issu de la palatalisation de /k, g/ + /æ, e, i/⁴ est déjà passé à /j̣C/ lors de la chute des voyelles finales atones, et ne provoque donc pas la rétention d'une finale atone, puisqu'il n'y a plus qu'une consonne simple (p.ex. PĀCEM → *pais* ; id. GGHF, p. 341) ; pour /k/ palatalisé, cette consonne doit d'ailleurs être /ts/ (palatalisation romane) ou /tʃ/ (palatalisation gallo-romane) et non /k/. Au vu de ces principes, le <e> des descendants de ÚSQUE ne devrait donc pas être de droit.

Fouché (1966, p. 914) propose, pour certaines de ces formes, une construction lexicalisée se terminant par AD, ce qui permet à l'étymon de se terminer par un /a/ qui justifie la rétention d'une voyelle finale atone : INDE USQUE AD → *†indyusqua*⁵ → *enjosque, enjusque* ; *†indesqua* → *desque* ; il ne propose en revanche pas de /a/ final pour *†ind(e)ūsque* → *jusque*.^{6,7} Si l'on acceptait cette réfection étymologique comme étant certaine et régulière, il conviendrait d'intégrer ces lemmes dans notre analyse des adverbes. Néanmoins, la distribution des défauts de <e> de ces quatre lemmes revêt un caractère exceptionnel qui ne plaide pas en faveur de leur intégration au sein du paradigme des adverbes où <e> est de droit : le défaut de <e> de ces quatre lemmes est strictement conditionné par l'initiale subséquente. FAGARD (2020, GGHF, p. 873) semble le sous-entendre en citant les formes *dusc'*, *dusk'* et *dusqu'*. L'usage de l'apostrophe, accompagné d'un exemple où *dusc'* se trouve devant initiale vocalique

alia : ANON. (1185*). *Roman d'Alexandre*. Ms. A ; Dees : Vendée, Deux-Sèvres ; compo. pic. et occ. orient., ca 1185 ; ms. poit., ca 1225 ; MS (ANON. 1937) ; DEAF : AlexArsL.

poit : ANON. (1217*). *Chronique traduite de l'Historia Karoli Magni et Rotholandi*. Dees : Vendée, Deux-Sèvres ; compo. saint., ca 1217 ; ms. saint., ca 1238 ; MS1 (AURACHER 1877) ; DEAF : Turpin¹A.

calen : CALENDRE (1215*). *Des empereors de Rome*. Dees : Haute-Marne ; compo. champ., 1215 ; ms. champ., ca 1235 ; MS2 (SETTEGAST 1879) ; DEAF : CaLendReS.

3: *Oltra* dans *calen* (p. 6) est le seul exemple, sur 20 occurrences, à ne pas provenir du Sud-Ouest.

SAMPSON (1980). « On the history of final vowels from Latin to Old French ».

4: /æ/ correspond au descendant du lat.class. A, devenu antérieur et donc agent de palatalisation (SÉGÉRAL et SCHEER 2020, GGHF, p. 212, 237).

Fouché (1966). *Phonétique historique du français — Les consonnes*.

5: Nous reproduisons la notation de Fouché.

6: Il importe de considérer que Fouché cite ces étymons pour parler d'un autre phénomène (le devenir du groupe /ndj/) ; il ne présente donc pas d'argument propre à expliquer la rétention d'une finale atone.

7: Pour toutes ces formes, le FEW et le REW s'en tiennent à ÚSQUE.

FAGARD (2020). « Prépositions et locutions prépositives ».

(7.2), incite à considérer que ces formes en défaut de <e> procèdent de l'élision.

- (7.2) Du plus vaillant dirai le some / qui fust d'Islande **dusc'a Rome**, / del bon conte Tibaut de Blois. (Gautier d'Arras, *Eracle*, 1176-1184, v. 51-53)
 « Je vais raconter l'histoire du plus vaillant de tous, d'Islande jusqu'à Rome : le bon comte Thibault de Blois. »
 (FAGARD 2020, p. 873 ; graisse de la source.)

C'est précisément cette distribution conditionnée par le contexte prévo-calique que l'on retrouve dans le NCA. Dans le corpus, l'ensemble de ces lemmes est attesté en tant qu'adverbes 1772 fois. Sur ces 1772 occurrences, 476 sont dotées d'un '-s adverbial' (*jusques*, *dusques*, etc.), et peuvent se trouver tant devant initiale vocalique que consonantique. Au sein des formes sans '-s adverbial', les formes dotées d'un <e> final semblent pouvoir être aussi bien situées devant initiale vocalique (230 occ.) que consonantique (165 occ.), mais une écrasante majorité des formes sans <e> se trouve devant une initiale vocalique (898 occ. sur 901 occ., soit 99,7%). Ces chiffres soulèvent plusieurs questions. D'une part, le contexte préconsonantique est en principe bien plus courant en a.fr. que le contexte prévo-calique ; ici, visiblement, les contraintes syntaxiques et sémantiques impliquées par les descendants de ÚSQUE inversent cette tendance en plaçant le mot devant une initiale vocalique six fois plus souvent que devant initiale consonantique.⁸ D'autre part, et comme ce sera le cas pour d'autres lemmes issus d'autres parties du discours, le défaut de <e> apparaît non seulement uniquement devant initiale vocalique, mais il est aussi largement dominant dans ce contexte (80% des occ. prévo-caliques sont en défaut de <e>). Or ce comportement très fortement conditionné par le contexte subséquent fait précisément défaut pour les adverbes que nous allons étudier ici, qui ne font pas montre d'une telle sensibilité au contexte subséquent (sec. 7.2.4, 7.3.2). L'absence de prédictibilité du <e> à partir de l'étymon latin et des règles de la phonétique historique ainsi que ce comportement idiosyncrasique pour un adverbe nous ont donc incité à maintenir ces lemmes hors des données qui sont détaillées dans ce chapitre.

7.1.2.2 *Ne* : motivation de l'exclusion

Un autre lemme, correspondant à l'adverbe de négation *ne* étiqueté en2_ne, a été exclu de nos analyses. Il n'a pas été exclu *a priori*, pour des raisons linguistiques, mais pour des raisons pratiques : sur l'ensemble du corpus, seules 60 formes lemmatisées en2_ne sont annotées pos = "ADV". Non seulement ce chiffre est déjà suspect en soi, étant bien trop faible, mais en plus il ne concerne que trois textes : une forme *ne* se trouve dans *sully2*, deux formes *ne* se trouvent dans *fabLesK*, et les 57 formes restantes, *ne* et *n*, se trouvent toutes dans *mir*. L'étiquette numérique de Dees, traduite en ADV pour l'attribut de pos, est 311. Pour toutes ces formes, Tree Tagger propose l'étiquette de partie du discours PRO_clit, mais sur l'ensemble du corpus, ce sont 43 404 formes *ne* et 22 526 formes *n* que Tree Tagger étiquette ainsi. Les autres formes *ne* et *n* rejoignent *ni* sous l'étiquette PROCON. L'ensemble des configurations d'association de ces attributs est donnée en table 7.1, où nous n'avons omis que les

8: Les contextes des occurrences avec '-s adverbial' sont comptés ici : le mot est placé devant initiale consonantique dans 239 occ., devant initiale vocalique dans 1531 occ., devant un mot commençant par un <h> dans 1 occ. et devant une ponctuation dans 1 occ.

sully2 : MAURICE DE SULLY (ca 1250). *Sermons de Maurice de Sully*. Compo. La Rochelle, ca 1250 ; ms. poit., ca 1250 ; MS2 (BOUCHERIE 1873), DEAF : *SermMaurpB*.

fabLesK : MARIE DE FRANCE (1180*). *Fables*. Ms. K ; Dees : nil ; compo. NO, ca 1180 ; ms. nil, ca 1310 ; MS (transcription inédite).

mir : GAUTIER DE COINCI (1224*). *Les miracles de Notre Dame de Soissons*. Ms. R ; Dees : Aisne ; compo. Soissonnais, ca 1224 ; ms. pic. mérid., ca 1310 ; CR (GAUTIER DE COINCI 1966) ; DEAF : *CoincyI10K*.

combinaisons dont la fréquence est inférieure à 10 occurrences et celles où la forme est annotée comme un nom.

TABLE 7.1 – Combinaisons des formes, parties du discours de Dees (deespos et, développées, pos) et de Tree Tagger (taggerpos) pour les formes *ne*, *ni* et *n* non nominales et dont F > 10.

Forme	deespos	pos	taggerpos	F
<i>ne</i>	319	PRO_clit	PRO_clit	38 869
<i>n</i>	319	PRO_clit	PRO_clit	21 448
<i>ne</i>	600	PROCON	PROCON	7247
<i>ne</i>	319	PRO_clit	PROCON	4681
<i>ne</i>	600	PROCON	PRO_clit	4478
<i>n</i>	600	PROCON	PRO_clit	1048
<i>n</i>	600	PROCON	PROCON	393
<i>ni</i>	600	PROCON	PROCON	248
<i>ne</i>	311	ADV	PRO_clit	39
<i>n</i>	319	PRO_clit	PROCON	39
<i>n</i>	311	ADV	PRO_clit	21

De fait, si l'annotation des formes *ne* et *n* représente une véritable difficulté pour Tree Tagger, les annotations de Dees ne semblent pas s'en sortir beaucoup mieux. D'une part, il faut considérer que, contrairement au fr.mod., l'a.fr. ne connaît pas la distinction systématique entre *ne* et *ni* (MÉNARD 1988 [1973], p. 198; BADIOU-MONFERRAN 2020, GGHF, p. 948), où seul *ni* représente la conjonction de coordination. Toute opération de lemmatisation doit donc se passer, pour l'a.fr., de cette distinction graphique. Par ailleurs, la caractérisation de la forme *ne* est problématique : elle représente la forme atone du négateur syntaxiquement autonome *non* ; cette forme atone apparaît dès le x^e s. et est qualifiée de « forme atone et conjointe au verbe » (MARCHELLO-NIZIA 1999b, p. 114), de « forme préverbale » (HANSEN 2020, GGHF, p. 1681-1682), ou de « forme non prédicative et non autonome » (COMBETTES et al. 2020, GGHF, p. 1252). Ces qualificatifs qui, en syntaxe comme en phonologie, font signe vers un processus de cliticisation, sont peu compatibles avec la nature syntaxique et phonologique des adverbes, et augmentent encore la complexité de l'annotation en parties du discours des formes *ne* et *n*. Ainsi, la majeure partie des formes *ne* et *n* se trouve étiquetée PRO_clit ou PROCON, tandis que l'annotation ADV est marginale (moins de 0,08% des formes *ne*, *ni*, *n*). Dans la mesure où ce qui correspondrait à l'adverbe négatif ne correspond pas, dans une écrasante majorité des cas, à l'attribut ADV, et où les occurrences de l'attribut ADV pour la négation sont limitées à trois textes, nous avons opté pour l'élimination de ce lemme, de manière à limiter sa capacité à parasiter les résultats des autres adverbes.

7.1.2.3 En somme

Sur les 81 lemmes détectés par Comparalem comme ayant une probable variation <e>/∅, nous en avons retenu 12. Les lemmes exclus l'ont été soit parce que leur <e> n'est pas de droit, soit parce que leur variation n'est pas une variation <e>/∅, puisqu'elle résulte d'ambiguïtés graphémiques (<e> pour /e/ ou <a> pour /ə/) ou d'erreurs d'annotation. Par ailleurs, nous avons exclu 20 occurrences en <a> dans des lemmes qui connaissent par ailleurs une véritable variation par défaut de <e> (« arière » : *arriera* (occ. exclues) ~ *arriere*, *arrier* (occ. conservées); « encore » : *encora* (occ.

MARCHELLO-NIZIA (1999). *Le français en diachronie : douze siècles d'évolution*.

HANSEN (2020). « La négation de proposition ».

COMBETTES et al. (2020). « Syntaxe de la phrase simple ».

exclues) ~ *encore, encor* (occ. conservées) « ensemble » : *ensembla* (occ. exclues) ~ *ensemble, ensembl* (occ. conservées)). Sans considération des contextes qui restreignent par la suite nos analyses dans Comparalem, ces 12 lemmes représentent 20 547 occurrences, dont 9333 (45%) sont en défaut de <e>. Enfin, comme indiqué en sec. suivante, ces 12 lemmes correspondent à 12 étiquettes d'attributs lemma mis en place par Tree Tagger ; après relemmatisation, ces 12 lemmes sont réduits au nombre de 9.

7.1.3 Relemmatisation

Après examen manuel et en contexte des occurrences, nous avons décidé de procéder à deux opérations de relemmatisation. Nous avons réuni sous le même nouveau lemme deux ensembles d'étiquettes de Tree Tagger : d'une part *derrier, derriere_derriere* et *derriere*, et d'autre part *or4* et *or4_or* (7.3). En effet, pour ces étiquettes, les occurrences annotées en tant qu'adverbes semblent bien représenter toujours le même lemme linguistique. La logique selon laquelle Tree Tagger a voulu distinguer ces lemmes ne nous apparaît pas évidente ; il a rangé des occurrences avec et sans <e> sous *derrier* comme sous *derriere*, et a créé une étiquette mixte *derrier_derriere* sous laquelle il range également les deux types de formes, et nous ne sommes pas parvenu à identifier des contextes syntaxiques qui auraient fait une différence. De même, les raisons de la division entre *or4_or* et *or* demeurent obscures ; on trouve des formes avec et sans <e> sous les deux étiquettes.

(7.3)	Tree Tagger	Nouveau lemme
	<i>derrier, derrier_derriere, derriere</i>	→ <i>derriere_relemma</i>
	<i>or4_or, or</i>	→ <i>or_relemma</i>

Au terme de cette opération de relemmatisation, il ne reste donc plus que 9 lemmes dans la liste de ceux que nous retenons pour le défaut de <e>.

7.2 Présentation des résultats

Notre corpus contient 170 479 adverbes situés dans l'un des contextes interrogés,⁹ qui correspondent à 711 lemmes. Une nouvelle fois, ce nombre est surestimé, comme le montre la redondance des étiquettes que nous avons fusionnées. Nous avons retenu 9 lemmes représentant une véritable alternance par défaut de <e> ; ces 9 lemmes représentent 14 538 occurrences, soit 8,5% des occurrences adverbiales considérées. Sur les 14 538 occurrences de ces 9 lemmes, 9977 sont en défaut de <e> ; ces 9977 occurrences défectueuses correspondent à un peu moins de 6% de toutes les occurrences adverbiales mais correspondent à 68% des occurrences de ces 9 lemmes. Ainsi, le taux de défaut de <e> au sein de l'ensemble de la catégorie est ici équivalent à celui des prépositions (PRÉP : 6%), mais l'absence de <e> pour les lemmes retenus est bien plus importante que pour les prépositions (PRÉP : 16%).

9: C'est-à-dire situés devant initiale vocale (sauf <i> et <u>) ou consonantique (sauf <v> et <h>), cf. sec. 3.1.1.

Ces données portent sur l'ensemble des neuf lemmes retenus. Or, l'un d'entre eux n'apparaît en défaut de <e> que dans un seul texte : seul

TABLE 7.2 – Quantité d’adverbes, de lemmes à variation, et d’occurrences en défaut de <e>.

	Nb.	Nb.var.	Taux	Nb.Var.Ø	Taux.Ø/Nb.
Lemmes	711	9	1,27%		
Occ.	170 479	14 538	8,53%	9977	5,79%
<i>Tx_Ø</i> (Nb.Var.Ø/Nb.var) :				67,94%	

Roland connaît des formes *ensembl*, le lemme conservant son <e> dans tous les autres textes (cf. sec. 7.2.3.1 *infra*). Si l’on considère que le défaut de <e> de « ensemble » est une forme d’idiosynchrisme de *Roland*, il n’y a pas lieu de le mélanger aux autres défauts, bien plus répandus. On peut donc envisager les données directement sans le lemme « ensemble », tel qu’en table 7.3.¹⁰ Si les valeurs des deux premières lignes du tableau évoluent logiquement peu, l’exclusion du lemme « ensemble » provoque en revanche une hausse sensible du taux de défaut sur les lemmes sujets au défaut (*Tx_Ø*), qui passe de 68 à 75%. En effet, aux 20 occurrences *ensembl* de *Roland* s’opposent 1600 occurrences du type sans défaut *ensemble* (avec variantes graphiques : *ensanle* etc.); le retrait de ces occurrences fait donc augmenter le taux de défaut.

10: Les valeurs de la colonne Nb. ne changent pas, dans la mesure où Comparalem les extrait du jeu de données brutes, avant toute opération (relemmatisation, exclusion, et donc, ici, exclusion de « ensemble »).

TABLE 7.3 – Quantité d’adverbes, de lemmes à variation, et d’occurrences en défaut de <e>, sans le lemme « ensemble ».

	Nb.	Nb.var.	Taux	Nb.Var.Ø	Taux.Ø/Nb.
Lemmes	711	8	1,13%		
Occ.	170 479	13 134	7,7%	9859	5,78%
<i>Tx_Ø</i> (Nb.Var.Ø/Nb.var) :				75,06%	

7.2.1 Typologie

Les neuf lemmes retenus sont donnés en (7.4). Comme indiqué par le classement présenté, quatre lemmes ont un dernier élément qui correspond au latin *HÓRAM*, et deux lemmes ont un dernier élément qui correspond au latin *RÉTRO*. Le lemme « desor », ici, ne correspond en revanche pas au latin *HÓRAM*, mais à *SÚPER* ou *SÚPRĀ*, dont il a été question au chapitre précédent (cf. sec. 6.2.1); il ne s’agit donc pas du *desor* temporel de *desormais* mais de l’adverbe de localisation spatiale.

- (7.4) (a) HÓRAM
 a) → « or »
 b) → « encore »
 c) → « lors »
 d) → « mar » (MÁLĀ HÓRĀ)
- (b) -RĒTRO
 a) → « ariere »
 b) → « derriere »
- (c) SÚPER → « desor »
 (d) ŪNQUAM → « onque »
 (e) INSĪMUL¹¹ → †*insēm*ul (REW, FEW) → « ensemble »

Sur ces 9 lemmes, seuls deux se terminent par un groupe consonantique (hors ajout d'un '-s adverbial') : « ensemble » et « onque ». Pour le premier, l'épenthèse de /b/ est indépendante du maintien de la finale atone : si les seules formes en défaut de <e> connaissent l'épenthèse consonantique (type *ensembl*), les formes épenthétiques sont aussi beaucoup plus courantes (type *ensemble*) que les formes sans épenthèse (*ensanle*), le type sans épenthèse ne représentant que 4% des occurrences.¹² L'attaque branchante /bl/ n'est par ailleurs pas nécessaire à la génération d'une finale atone ; /ml/ ne fait pas partie des groupes licites en position finale listés par SAMPSON (1980) (ce qui impose donc une voyelle de soutien), et l'auteur propose une métathèse †*insēm*ul → proto-a.fr. [ensem^lu] (p. 39). Pour les autres lemmes, le schwa susceptible de chuter est précédé d'une consonne simple, déjà simple en latin (HÓRAM) ou qui fait suite à la simplification d'un groupe *obstruante + liquide* (/pr/ dans SÚPER après la chute de la posttonique, /tr/ dans -RĒTRO).

Si nos 9 lemmes ne représentent que 1,27% des 711 lemmes adverbiaux dotés d'occurrences dans les contextes interrogés par Comparalem, il convient de noter que seuls 62 lemmes se terminent par un <e> qui ne soit pas partie prenante d'une finale <que> (exclus pour éviter l'influence analogique de ŪNQUAM et les lemmes formés sur ÚSQUE). Sur ces 62 lemmes, seuls 12 ont une fréquence supérieure à 100 occurrences, et 4 de nos 9 lemmes appartiennent à ces 12 lemmes, comme indiqué en table marginale 7.4.

Si cette liste n'est évidemment pas complète, dans la mesure où elle n'interroge que les lemmes et non leurs formes (manquant, ainsi, nos lemmes « or », « lors » et « mar »), elle donne une idée du très faible nombre de lemmes adverbiaux qui sont à la fois en situation de défaut de <e> (c'est-à-dire qu'ils disposent d'un <e> de droit) et en quantité suffisante pour exhiber ce défaut de manière significative. Aussi, il convient de considérer qu'approximativement la moitié des lemmes qui pourraient avoir un défaut de <e> ont, en effet, des occurrences défectueuses.

7.2.2 Distribution diachronique générale

Les figures 7.1 *supra* et 7.2 et 7.3 en pages suivantes présentent trois vues différentes de l'évolution diachronique générale des adverbes retenus. En fig. 7.1, les 9 lemmes sont comptabilisés, et les formes dotées d'un '-s adverbial' suivant un <e> (p.ex. *unkes*, *ores*) sont comptées comme des formes sans défaut de schwa. La situation représentée est fort hétérogène,

11: La descendance impose de considérer que le préfixe *in-* ne compte pas dans le domaine accentuel.

12: L'étendue géographique du type non épenthétique de « ensemble » est singulièrement circonscrite au département du Nord et, dans une moindre mesure, aux régions Somme, Pas-de-Calais et Oise ; cf. carte 488 dans DEES (1987, p. 488).

TABLE 7.4 – Fréquence adverbiale des lemmes se terminant en <e> (hors <que>). Il ne s'agit pas, ici, des données filtrées par les contextes de Comparalem, mais d'une interrogation sur l'ensemble du corpus.

lemma	nb.occ.
mie	4262
encore	2055
ensemble	1621
ariere	1170
le	568
mèisme	427
outrel_outre	368
derriere ^a	314
prime	283
gaire	108
sempr	107
estrel_estre	100

^a. Ici, la fréquence correspond à notre relemmatisation, c'est-à-dire aux occurrences de *derrier*, *derriere* et *derrier_derriere*.

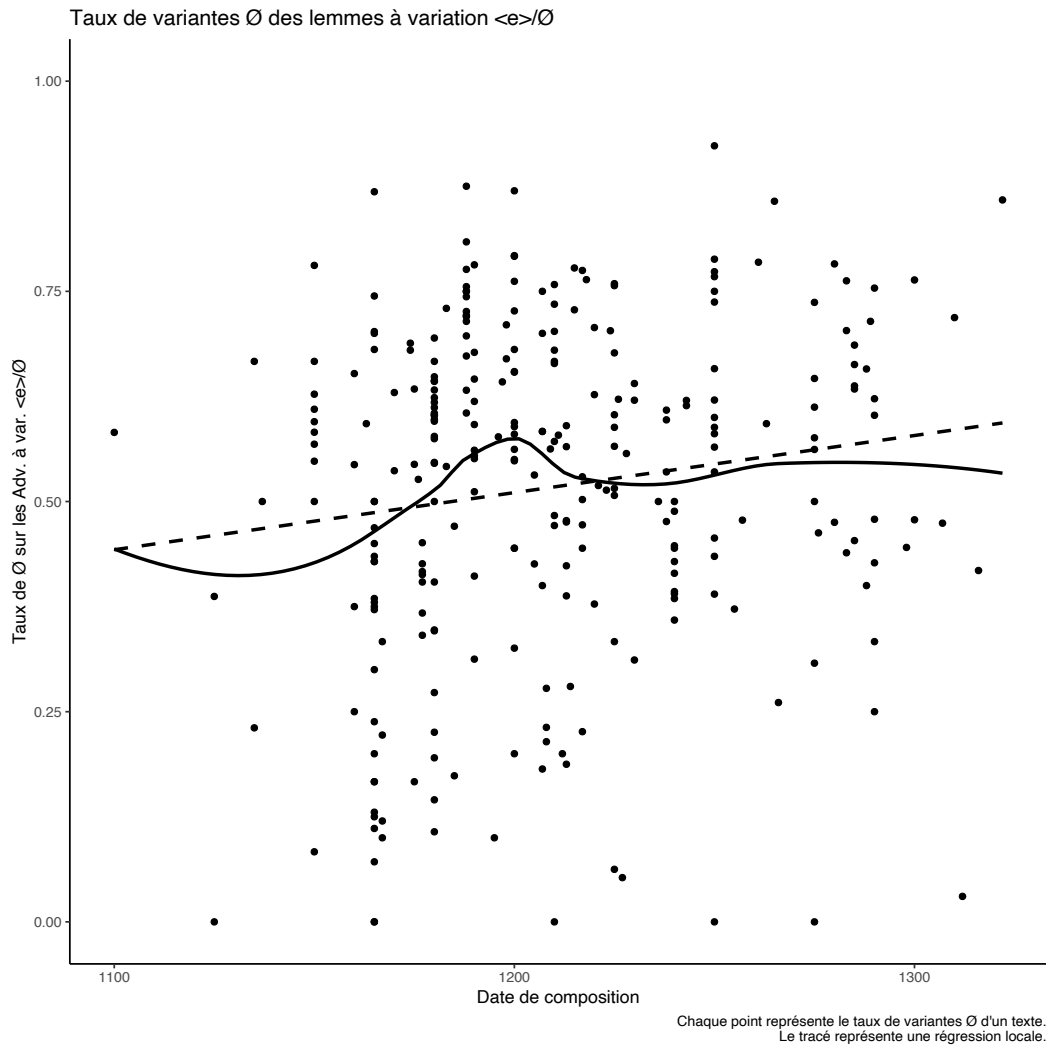


FIGURE 7.1 – Distribution diachronique des variantes d’adverbes en défaut de <e>.

les données s’étendant de 0 à 92% de défauts de <e> par texte, pour une moyenne de 52%. Tel que le figure la régression linéaire en pointillés, la tendance générale est à l’augmentation des taux de défaut de <e> alors que les textes se font plus tardifs. La régression locale, sensible aux fluctuations temporaires des taux, montre un taux moyen légèrement plus élevé autour de 1200; pour les adjectifs, un pic analogue se situait dans la seconde moitié du ^{xii}e s. (sec. 5.2.2), tandis que pour les prépositions le pic, ou plateau, visible uniquement devant initiale vocalique, se situait dans la seconde moitié du ^{xiii}e s. (sec. 6.2.4).

Une telle présentation, si elle a l’avantage de prendre en compte toutes les formes adverbiales de nos 9 lemmes, intègre un facteur d’interférence : la présence des ‘-s adverbiaux’. Ceux-ci font interférence dans la mesure où le schwa qui les précède n’est plus en position finale absolue. Dans les sections *infra*, pour le besoin des analyses en fonction du contexte subséquent, ces formes sont omises. Leur omission provoque naturellement une remontée des taux de défaut de <e>, puisque ceux-ci ne sont plus calculés que sur l’assiette des formes à <e> final et des formes sans

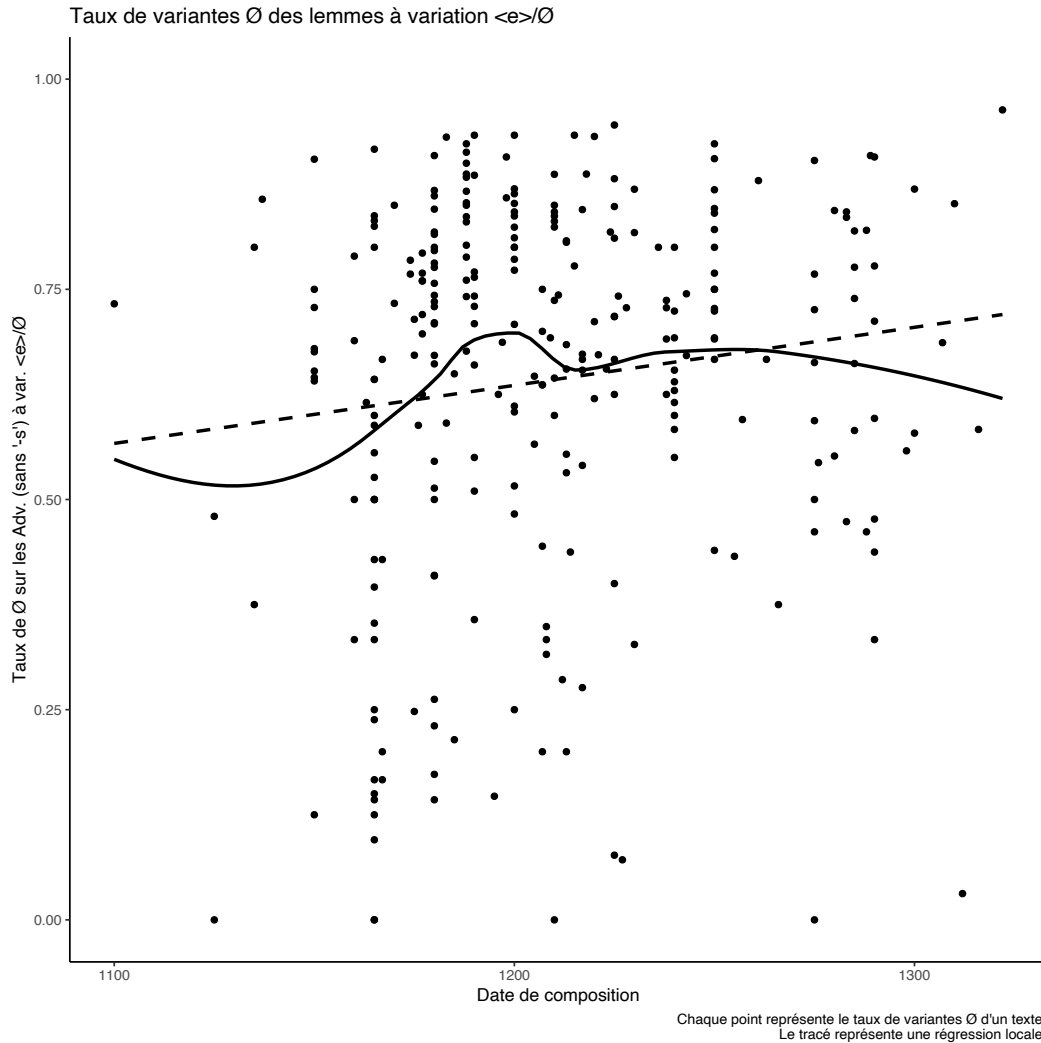


FIGURE 7.2 – Distribution diachronique des variantes d'adverbes en défaut de <e>, à l'omission des formes dotées d'un *-s* adverbial.

<e> final. Ainsi, en fig. 7.2, si l'on retrouve toujours la même tendance à l'augmentation des taux de défaut de <e>, la moyenne de ces taux est bien plus élevée, s'établissant à 64% (contre 52% en fig. 7.1). La même concentration de taux élevés autour de 1200 est également présente, avec, évidemment, un sommet plus élevé.

Enfin, et comme cela sera développé *infra* (sec. 7.2.3.1), le lemme « ensemble » ne connaît guère de défauts de <e> que dans *Roland*. Si l'on admet qu'il s'agit-là d'une idiosyncrasie de *Roland*, ou d'un archaïsme qui disparaît dès le début de notre corpus, il convient d'observer la situation diachronique sans ce lemme, pour empêcher que toutes les attestations sans défaut de <e> de « ensemble » n'écrasent les taux des autres lemmes. La diachronie du jeu de données sans *'-s adverbial'* et sans le lemme « ensemble » est représentée en fig. 7.3. Les deux tendances observées sur les deux figures précédentes s'y retrouvent encore une fois (croissance générale des taux de défaut et pic autour de 1200), avec des valeurs plus élevées, le taux moyen de défaut par texte s'élevant à 71%.

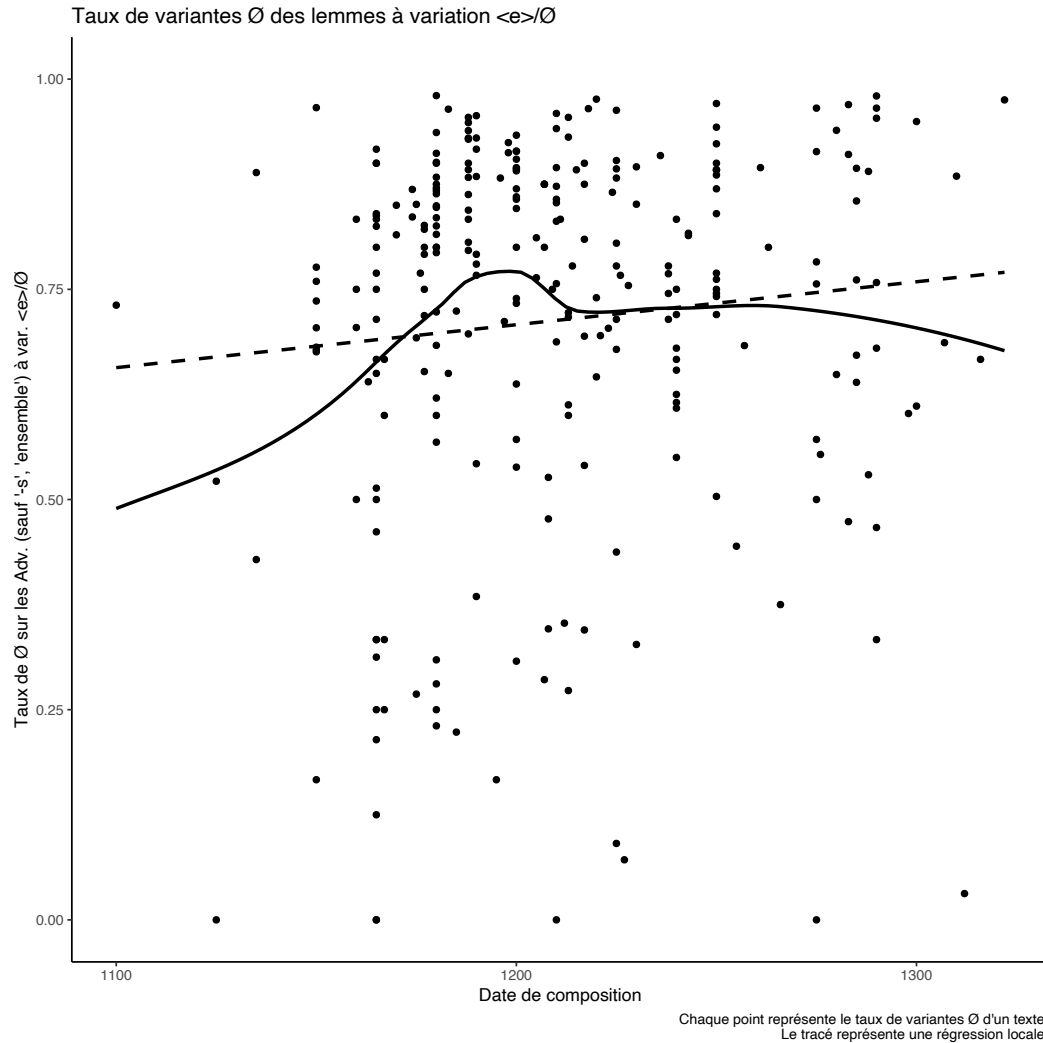


FIGURE 7.3 – Distribution diachronique des variantes d'adverbes en défaut de <e>, à l'omission des formes dotées d'un -s adverbial et du lemme « ensemble ».

Si cette tendance à la croissance générale des défauts de <e> n'est liée ni aux formes dotées d'un '-s adverbial' ni au lemme « ensemble », elle n'est cependant pas fiable. Nous démontrerons *infra* (sec. 7.3.2 et 7.3.4) qu'il s'agit d'un artefact, principalement dû aux textes anglo-normands, dont les taux de défaut sont à la fois bas et précoces. De fait, cette tendance à la hausse qu'on observe sur l'ensemble du corpus n'est observable ni en Angleterre ni dans le domaine continental, et résulte uniquement de la fusion de ces deux domaines linguistiques.

7.2.3 Distribution des lemmes en défaut de <e> : diachronie et fréquence

Les graphiques en fig. 7.4 et 7.5 en pages suivantes représentent, pour chaque texte, le taux de défaut de <e> de chaque lemme, devant initiale vocalique en fig. 7.4, et devant initiale consonantique en fig. 7.5. Ces graphiques montrent que, contrairement au nivellement induit par le

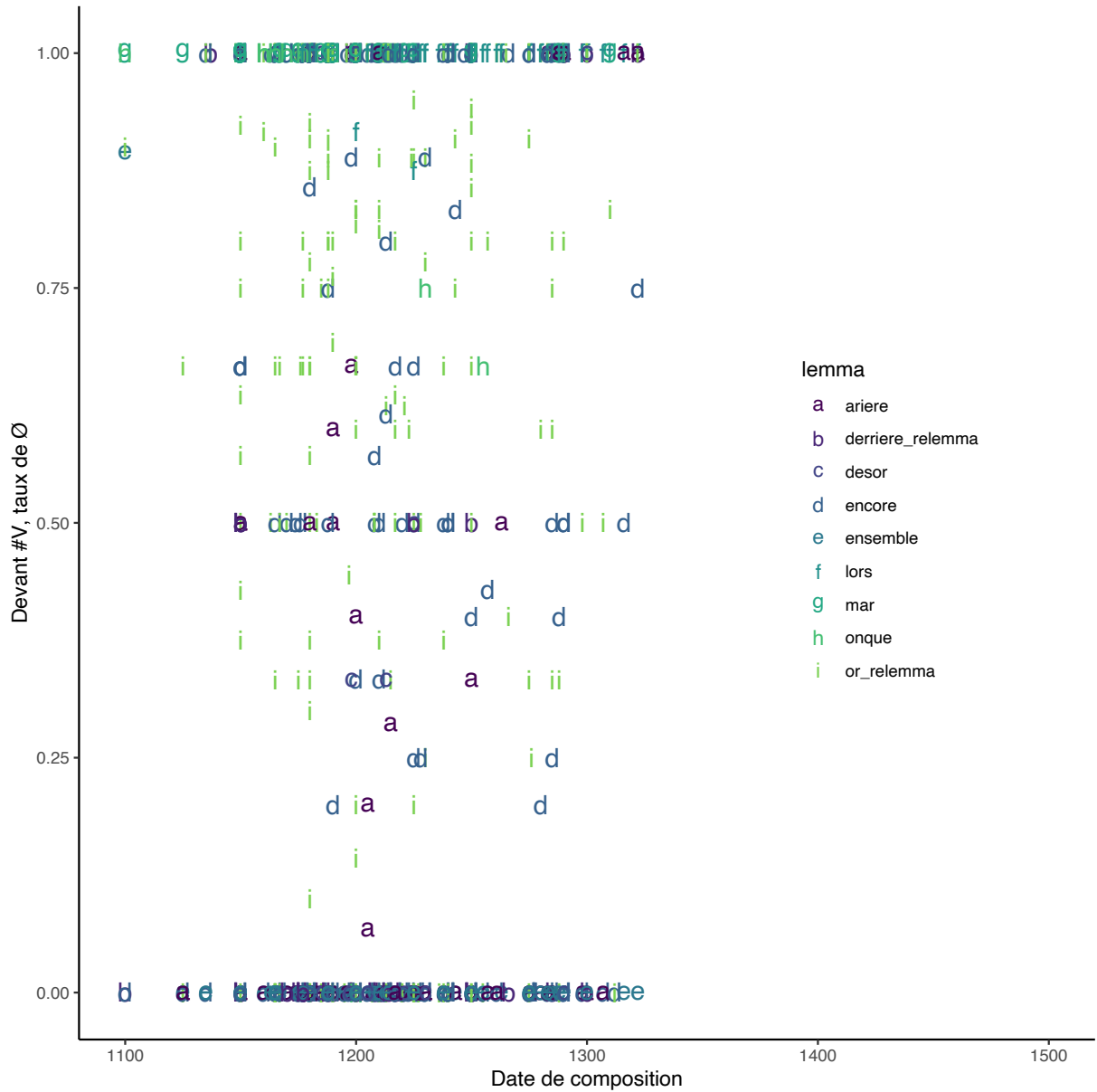
calcul du taux par texte (tous lemmes compris) utilisé en fig. 7.1, 7.2 et 7.3 *supra*, les lemmes ont tendance à se concentrer soit à 0% soit à 100%, ce qui, dans bien des cas, s'explique par une seule apparition du lemme dans le texte. De même, les points qui sont alignés sur 50% témoignent, pour la plupart, de textes qui n'ont que deux occurrences des lemmes en question. La variation plus subtile, qui concerne les lemmes présents plus de deux fois par texte, ne concerne donc qu'un ensemble réduit de lemmes.

Cette variation plus subtile concerne, au premier chef, le lemme « or » (fr.mod. « maintenant »). Celui-ci, par sa fréquence élevée (6027 occ.), permet des taux qui s'affranchissent des paliers que représentent les taux de 0%, 25%, 33%, 50%, 66%, 75% et 100% pour les lemmes attestés moins de 6 fois par texte. La répartition des points représentant ce lemme, qui sont en forme de 'i' sur les fig. 7.4 et 7.5 couvre par ailleurs l'entièreté de l'échelle des ordonnées, de 0% à 100%, et les taux moyens de 74% devant initiale vocalique et 79% devant initiale consonantique n'impliquent donc pas l'absence de taux inférieurs à 50%.

D'autres lemmes sont concernés par un fort degré de variation : « encore » (55% de défaut __#V ; 58% __#C), « desor » (31% __#V ; 80% __#C) et « derriere » (23% __#V ; 35% __#C) ont des taux moyens pour au moins un contexte compris entre 25 et 75% de défauts de <e>. À l'opposé du spectre, d'autres lemmes connaissent une variation marginale, soit par conservation dominante du <e> (« ensemble » : 3% __#V ; 0% __#C), soit par chute presque systématique du <e> (« mar » : 100% __#V ; 95% __#C ; « lors » 100% __#{V, C}). Dans ce qui suit, et comme pour les autres parties du discours, nous détaillons l'apparition des premières occurrences défectueuses des lemmes, ici concentrées dans *Roland* et dans la première moitié du XII^e s.

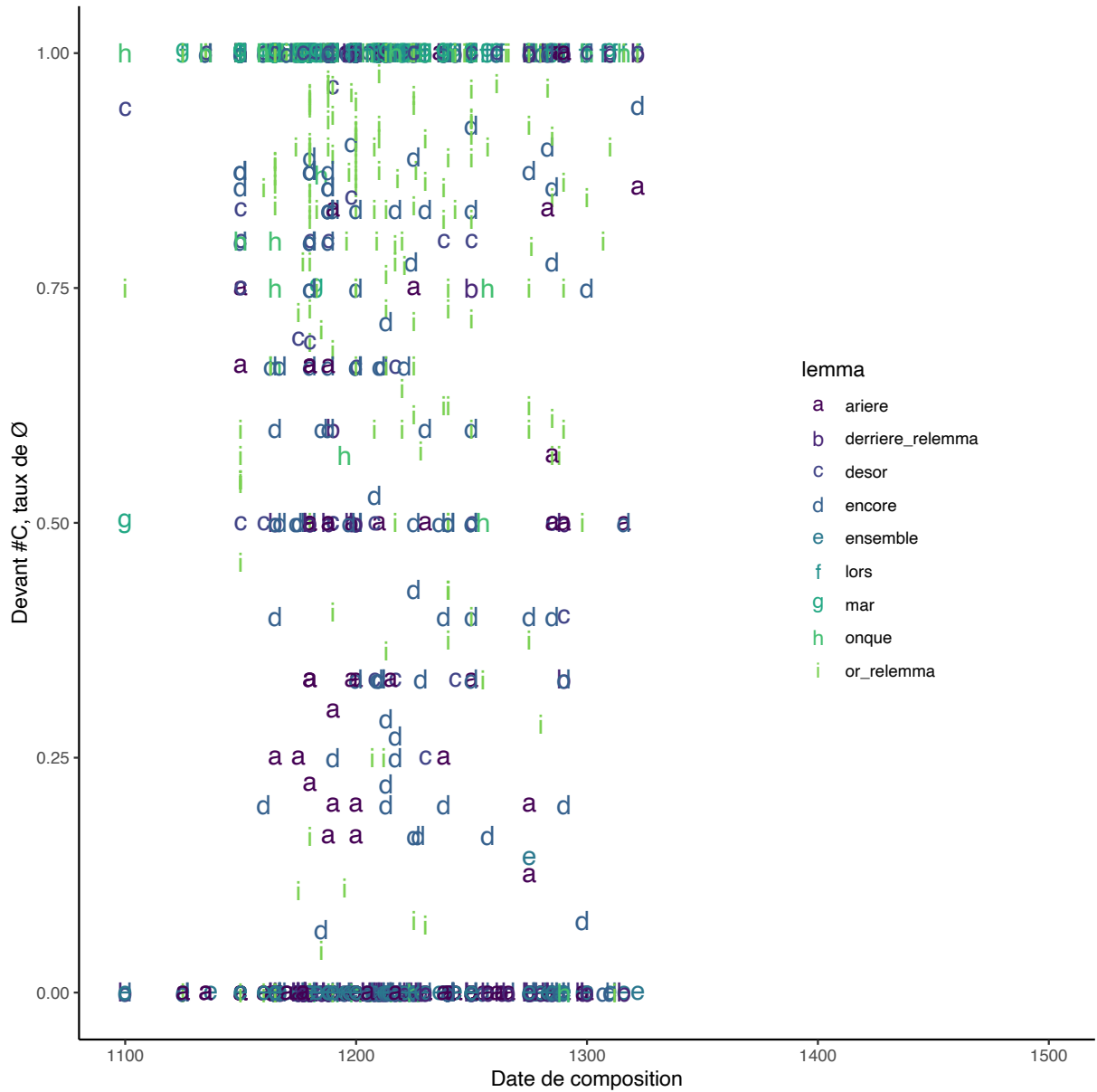
7.2.3.1 Lemmes attestés en défaut de <e> dès *Roland*

Cinq lemmes sont attestés en défaut de <e> dès *Roland* : « ensemble », « desor », « or », « mar » et « onque ». *Ensembl* n'est attesté que devant initiale vocalique, avec 17 occurrences (7.5), tandis que *desor* n'est attesté que devant initiale consonantique, avec 16 occurrences (7.6). Les autres lemmes sont attestés dans les deux contextes, dans des proportions variables : *or* est attesté 9 fois devant initiale vocalique (7.7) et 15 fois devant initiale consonantique (7.8), *mar* est attesté 2 fois devant initiale vocalique (7.9) et 8 fois devant initiale consonantique (7.10), et *unc* est attesté 1 fois devant initiale vocalique (7.11) et 4 fois devant initiale consonantique (7.12).



lemma	nb.e	nb.Ø	tot.	tx.Ø
mar	0	66	66	1.00
lors	2	592	594	1.00
onque	3	38	41	0.93
or_relemma	474	1347	1821	0.74
encore	220	264	484	0.55
desor	42	19	61	0.31
derriere_relemma	50	15	65	0.23
ariere	271	46	317	0.15
ensemble	585	17	602	0.03

FIGURE 7.4 – Distribution diachronique des alternances <e>/<Ø> selon les lemmes, adverbes suivis d’une initiale vocalique.



lemma	nb.e	nb.Ø	tot.	tx.Ø
lors	0	2275	2275	1.00
mar	9	188	197	0.95
onque	62	449	511	0.88
desor	97	382	479	0.80
or_relemma	901	3305	4206	0.79
encore	505	702	1207	0.58
derriere_relemma	125	66	191	0.35
ariere	514	105	619	0.17
ensemble	801	1	802	0.00

FIGURE 7.5 – Distribution diachronique des alternances <e>/<Ø> selon les lemmes, adverbes suivis d’une initiale consonantique.

- (7.5) *ensembl od lui rollant e oliver*
« avec lui sont Roland et Olivier »
(rolandox, v. 104, trad. SHORT 1990)
- (7.6) *un faldestoed i unt mis d olifan*
desur s asiet li païen baligant
« [ils] y installent un trône d'ivoire
Là le païen Baligant prend place » [= il s'assiet dessus]
(rolandox, v. 2654, trad. *id.*)
- (7.7) *dist bramimunde 'or oi mult grant folie*¹³
« Bramimonde dit : « voilà des paroles de fous ! »
(rolandox, v. 2714, trad. *id.*)
- (7.8) *respunt marsilie 'or diet nus l orrum*
« Marsile répond : « qu'il parle ; nous l'écouterons ! »
(rolandox, v. 424, trad. *id.*)
- (7.9) *respudent franc 'ja mar en vivrat uns*¹⁴
« Les Francs répondent : « On aura tord d'en laisser vivre un seul »
(rolandox, v. 3951, trad. *id.*)
- (7.10) *il dist al rei 'ja mar crerez marsilie*
« Il dit au roi : « N'allez surtout pas croire Marsile ! »
(rolandox, v. 196, trad. *id.*)
- (7.11) *unc einz ne puis ne fut si fort ajustee*
« depuis ce temps, ni avant, il n'y eut pareil affrontement »
(rolandox, v. 3394, trad. *id.*)
- (7.12) *unc mais nuls hom en tere n en vit plus*
« nul être humain n'en vit jamais d'avantage »
(rolandox, v. 1039, trad. *id.*)

13: *Or* prend ici déjà son sens de renforcement de l'exclamation (DMF « or2 » B.3). La lecture littérale *or oi* « j'entends maintenant » ne serait pas adaptée à ces contextes.

14: *mar*, renforcé par *ja*, renvoie aux conséquences funestes (AND : « to one's misfortune, disastrously ») ou exhorte directement à les éviter (AND : « never ever, under no circumstance ; God forbid »), c'est le deuxième sens qui est en jeu ici, comme en (7.10).

Pour tous ces lemmes, le texte dispose aussi d'occurrences sans défaut de <e> : *ensembl*, *desure*, *ore*, *mare* ; toutes les occurrences de « onque » sans défaut de <e> sont pourvues d'un '-s adverbial' (*unkes*, *unches*) et sont donc exclues de notre jeu de données. Les proportions sont variables selon les lemmes, avec 5 occurrences *ensembl* (3 __#C et 2 __#V), 1 occurrence *desure* (__#C), 6 occurrences *ore* (5 __#C et 1 __#V), 8 occurrences *mare* (__#C) et, hors des données que nous entrons dans Comparalem, 30 occurrences *unkes* ou *unches*.

Spécificité de *ensembl*. Les 17 attestations *ensembl* de *Roland* sont singulières. Non seulement ce sont les seules à n'intervenir que devant initiale vocalique (ce qui est analogue au comportement du <e> des prépositions « entre », « contre » et « encontre » ; cf. sec. 6.2.4.1), mais elles représentent par ailleurs les seules attestations du défaut de <e> de ce lemme devant initiale vocalique dans le corpus ; une attestation devant initiale consonantique apparaît plus tard, dans *darm*. C'est pour cette raison que ce lemme est par la suite exclu de la plupart des analyses que nous menons avec Comparalem, dans la mesure où il n'est pas possible de déterminer s'il s'agit d'une idiosyncrasie de *Roland* ou d'un trait archaïque. Enfin, ces occurrences *ensembl* sont toujours en collocation avec *od*, préposition (« avec ») issue de APUD (cf. sec. 7.2.3.4).

darm : ANON. (1275*). *Médecinaire Liègeois du XIII^e siècle*. Dees : Wallonie ; compo. liég., ca 1275 ; ms. wall, ca 1275 ; CRI (HAUST 1941) ; DEAF : MédLiégH.

Annotation en POS : adverbe ou préposition ? Si le défaut de <e> de « desor » prépositionnel (ou du moins annoté en tant que tel) n'apparaît

que *ca* 1165 avec *lanvals* (sec. 6.2.3.2, p. 196), on voit ici que la forme sans <e> apparaît, en position adverbiale, dès le premier texte du corpus. Au-delà de l'ambiguïté de certains items qui, en a.fr., fonctionnent aussi bien comme adverbes que comme prépositions (PICOCHÉ et MARCHELLO-NIZIA 1994 [1989], p. 271; MARCHELLO-NIZIA 1999b, p. 120; MÉNARD 1988 [1973], p. 262), il y a lieu de penser ici que la décision de Tree Tagger d'annoter toutes les occurrences « desur » de *Roland* comme étant des adverbes est fautive. En effet, certaines occurrences *verbe + desur + syntagme nominal* orientent clairement l'analyse vers la constitution d'un syntagme prépositionnel, alors que Tree Tagger a annoté ces occurrences du lemme comme adverbiales. Ainsi, dans l'exemple en (7.13), il n'y a pas lieu de considérer que *desur* est adverbial, et l'édition numérique de la BFM l'annote d'ailleurs comme une préposition.

(7.13) *oliver est desur*¹⁵ *un pui haut muntez*
« Olivier est monté sur une hauteur »¹⁶
(rolandox, v. 1017, trad. pers.)

7.2.3.2 Lemmes attestés en défaut de <e> au cours du XII^e s.

Les quatre autres lemmes retenus sont attestés en défaut de <e> dès la première moitié du XII^e s., entre *ca* 1213 (« lors ») et *ca* 1150.

Lors. Le lemme « lors » apparaît en défaut de <e> dès *gorm* (7.14), le troisième texte de notre corpus, lui aussi anglo-normand.¹⁷ Ce texte contient également des occurrences en défaut de <e> *mar* et *or* et des occurrences non défectueuses *ore*, *ensemble*, *ariere* et *uncore*. Aussi, dans *gorm*, seul « or » présente des formes avec et sans défaut de <e>, en l'occurrence une forme *ore* (devant initiale vocalique) pour sept formes *or* (2_#V et 5_#C).

(7.14) *si lors ne vait a lui juster*
« si, à ce moment, il n'allait pas le rejoindre »
(gorm, p. 3, trad. pers.)

Encor. Le lemme « encore » apparaît en défaut de <e> dès le texte suivant, avec deux occurrences dans *mrgrî*, devant initiale consonantique (7.15) comme devant initiale vocalique (7.16). Ce texte contient également deux occurrence *desur*, deux occurrences *onc* et trois occurrence *or*, et seulement deux occurrences de nos 10 lemmes sans défaut de <e>, avec *ensemble* et *ariere*. Si l'on admet que le défaut de <e> sur le lemme « ensemble » est, en définitive, une caractéristique propre à *Roland*, ce texte ne contient donc qu'une seule occurrence non défectueuse, contre neuf occurrences défectueuses.

lanvals : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. S; Dees : nil; compo. N.-O., ca 1165; ms. frc., ca 1300; MS2 (MARIE DE FRANCE 1958); DEAF : MarieLanvR².

PICOCHÉ et MARCHELLO-NIZIA (1994 [1989]). *Histoire de la langue française*.

MARCHELLO-NIZIA (1999b). *Le français en diachronie : douze siècles d'évolution*.

MÉNARD (1988 [1973]). *Syntaxe de l'ancien français*.

15: NCA : pos = "ADV"

BFM : fropos = "PRE"

16: Le vers est par ailleurs défectueux; la BFM édite *Oliver est desur un pui...* (avec un balisage <gap> dans le fichier XML), SHORT (1990) édite *Oliver muntez desur un pui halçur* (probablement la leçon d'un autre ms.); le vers du NCA est hypermétrique (il fait 11 syllabes métriques).

gorm : ANON. (1125*). *Gormont et Isebart*. Dees : Angleterre; compo. frc., ca 1125; ms. a.n., ca 1213; MS2 (BAYOT 1931); DEAF : GormB.

17: La localisation philologique donne 'francien' pour la composition comme pour la copie, mais Dees calcule une localisation en Angleterre avec un bon degré de certitude (coef. dial. 80).

mrgrî : WACE (1135*). *Vie de Sainte Marguerite*. Ms. M; Dees : [Haute-Marne]; compo. norm., ca 1135; ms. tour., ca 1238; CR3 (WACE 1932b); DEAF : WaceMargmF.

- (7.15) *n en avoit pas encor tot dit quant le solail resusir vit*
« elle n'avait pas encore raconté toute l'histoire quand elle vit le soleil se lever »
(mrgri, p. 2, trad. pers.)
- (7.16) *encor estoit en oreisons la sainte virgene a genoillons quant*
« la sainte vierge était encore en prière, à genoux, quand... »
(mrgri, p. 7, trad. id.)

Parallèlement, l'autre version de la *Vie de sainte Marguerite* de Wace du corpus, *marga*, ne contient que deux occurrences défectueuses (*encor* et *or*, devant initiale vocalique) pour quatre occurrences non défectueuses (1 occ. *arriere* devant initiale consonantique, 1 occ. *ensamble* et 2 occ. *encore* devant initiale vocalique). L'une de ces occurrences *encore* (7.17) correspond au contexte du second *encor* de *mrgri* (7.16), ce qui montre probablement une incertitude sur le traitement du <e> de l'adverbe dans le processus de copie.

- (7.17) *encore estoit en orisons la sainte virge a jenoillons la sainte virge a jenoillons quant*¹⁸
(marga, p. 12)

marga : WACE (1135*). *La vie de Sainte Marguerite de Wace*. Ms. M; Dees : Haute-Marne; compo. norm., ca 1135; ms. art., 1267; CR3 (WACE 1932a); DEAF : [WaceMargaF](#).

18: L'origine de la répétition de *la sainte virge a jenoillons* ne nous est pas connue.

Derrier. Le lemme « derriere » apparaît en défaut de <e> ca deux ans plus tard, dans *juise2* (7.18). Ce texte contient également 8 occurrences *or* devant initiale vocalique et 5 devant initiale consonantique, et ne contient aucune occurrence sans défaut de <e> de nos lemmes dans les contextes interrogés par nos requêtes. Le texte contient en revanche trois formes suivies d'une initiale <i> ou <v> (2 occ. *or vos* et 1 occ. *or irai*),¹⁹ ce qui fait que son taux de défaut de <e> n'est pas de 100% mais de 86%, tel qu'il est représenté en fig. 7.2 *supra*.

- (7.18) *et derrier et devant*
« de tous côtés »
(juise, p. 2, trad. pers.)

juise2 : ANON. (1137*). *Li ver del Juise*. Dees : Wallonie; compo. liég., ca 1137; ms. wall., ca 1210; CR2 (RANKKA 1982); DEAF : [JuiseR](#).

19: Ces occurrences n'appartiennent pas aux données contextuelles traitées par Comparalem (cf. sec. 3.1.1), mais appartiennent bien aux données non contextuelles telles qu'en fig. 7.1, 7.2 et 7.3.

Arrier. Notre dernier lemme à être attesté en défaut de <e> est l'autre lemme formé sur -RĒTRO, « ariere », qui apparaît en défaut de <e> au milieu du XII^e s., avec l'occurrence *arrier* de *nima2*, *nima3* et *nima4* (7.19).

- (7.19) *arrier enmi la sale*
« au fond, au milieu de la salle »
(nima2, p. 19; nima3, p. 12; nima4, p. 18; trad. pers.)

Ces textes, comme les autres versions du *Charroi de Nîmes*, contiennent de nombreuses occurrences défectueuses des lemmes précédemment cités (sur l'ensemble de la famille textuelle, 351 occ. défectueuses), mais ils comportent également des occurrences non défectueuses (sur l'ensemble de la famille textuelle, 137 occ. non défectueuses). Si tous les textes du *Charroi de Nîmes* ne contiennent pas un syntagme équivalent à (7.19), *nimb1* et *nimb2* ont en revanche un syntagme très proche, remplaçant simplement *arrier* par *par derrier*, dans une occurrence également défectueuse, donc.

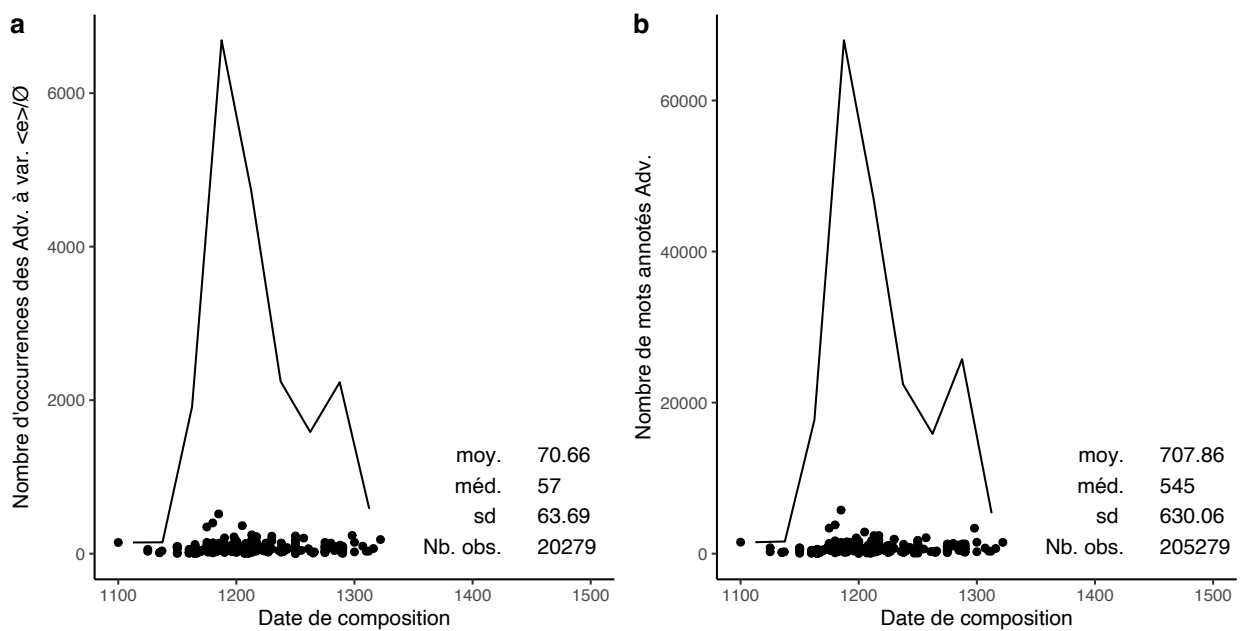
nima2 : ANON. (1150*). *Le charroi de Nîmes*. Ms. A2; Dees : Nièvre, Allier; compo. nil, ca 1150; ms. frc., ca 1263; MS1 (transcr. inédite L. Schoesler).

nima3 : ANON. (1150*). *Le charroi de Nîmes*. Ms. A3; Dees : Nièvre, Allier; compo. nil, ca 1150; ms. lorr., ca 1325; MS1 (transcr. inédite L. Schoesler)

nima4 : ANON. (1150*). *Le charroi de Nîmes*. Ms. A4; Dees : Nièvre, Allier; compo. nil, ca 1150; ms. frc., ca 1283; MS1 (transcr. inédite L. Schoesler).

— *Spécificité de « derriere » et « arriere ».* Les deux lemmes formés sur -RÉTRO ne sont pas seulement les deux lemmes à apparaître en dernier en défaut de <e> dans notre corpus, ce sont aussi les deux seuls lemmes dont le taux de défaut de <e> est inférieur à 50% (cf. table 7.5 *infra*), avec un taux de 16% pour « ariere » et de 32% pour « derriere ». S'il n'est pas directement possible de lier cette spécificité à leur étymon dans la mesure où l'étymon SŪPER de *desor* comprend le même contexte *obstruante + liquide* réduit à la seule obstruante, la conjonction de leurs faibles taux de défaut de <e> et de leur apparition légèrement plus tardive reste remarquable. Ce n'est pas non plus par l'effet d'une absence ou rareté des attestations que ces lemmes sont tardifs; dans *Roland*, par exemple, on trouve déjà six occurrences des lemmes (4 occ. *derere*, 1 occ. *darere* et 1 occ. *arere*).

En somme. Non seulement nos 9 lemmes sont attestés en défaut de <e> dans la première moitié du XII^e s., mais cinq d'entre eux le sont dès *Roland*. Cette fois-ci, et contrairement à ce que nous avons vu pour les catégories du discours précédentes, tous les lemmes sont donc attestés avant l'explosion quantitative du corpus, dans la seconde moitié du siècle. Cette augmentation substantielle de la taille du corpus est représentée en fig. 7.6, avec en (a) le nombre absolu d'occurrences adverbiales de nos 9 lemmes et en (b) le nombre absolu d'occurrences adverbiales pour tous les lemmes; sur ces figures, la courbe représente la somme du nombre d'occurrences des textes sur une fenêtre de 25 ans. Il ne fait donc pas de doute, ici, que la variation <e>/∅ de ces adverbes est antérieure au corpus, et peu sensible à un filtre lié à la quantité de données disponibles.



Chaque point représente un texte.

Les courbes relient la somme des occurrences sur une fenêtre de 25 ans.

FIGURE 7.6 – Distribution diachronique des occurrences des lemmes d'adverbes à défaut de <e> (a), en regard de distribution diachronique des occurrences de tous les lemmes d'adverbes.

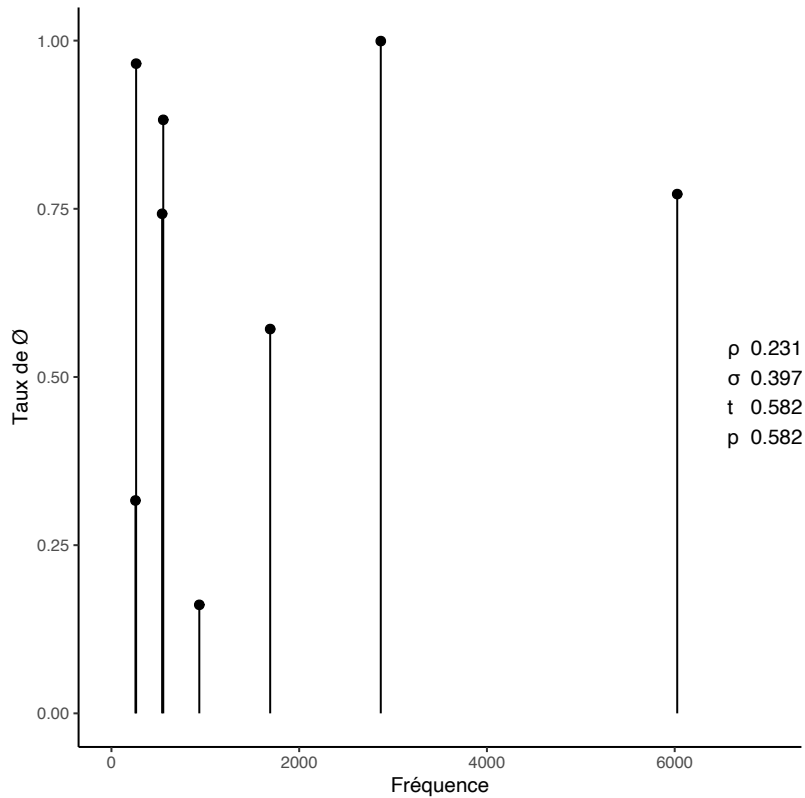


FIGURE 7.7 – Distribution des lemmes en fonction de leur fréquence et de leurs taux de défaut de <e>, adverbes, à l’exception des formes dotées d’un ‘-s adverbial’ et du lemme « ensemble ».

TABLE 7.5 – Relation entre fréquence et taux de défaut de <e>, pour chaque lemme d’adverbe.

lemma	nb.Ø	nb.E	frq.	tx.Ø
or_relemma	4652	1375	6027	77%
lors	2867	2	2869	99,9%
encore	966	725	1691	57%
arriere	151	785	936	16%
onque	487	65	552	88%
desor	401	139	540	74%
mar	254	9	263	97%
derriere_relemma	81	175	256	32%

7.2.3.3 Fréquence des lemmes et défauts de <e>

La fig. 7.7 représente la relation entre les taux de défaut de <e> des lemmes et leur fréquence d’emploi. Les données, ici, sont d’emblée considérées à l’exception des occurrences se terminant par <es> pour schwa suivi d’un ‘-s adverbial’ et du lemme *ensemble*, restreint à *Roland*. Les données sur lesquelles s’appuie cette figure sont reproduites en table 7.5, qui permet d’identifier les lemmes représentés par les lignes du graphique. Si le faible nombre de lemmes ne permet pas une analyse très assurée, ces données semblent invalider toute suspicion d’une corrélation

entre la fréquence lexicale et les taux de <e>. En effet, les indicateurs d'une corrélation qui serait linéaire sont très insatisfaisants (Pearson : $\rho = 0,2$; $p = 0,58$), et une simple inspection visuelle du graphique contredit également l'existence d'une corrélation qui ne serait pas linéaire, à la manière de l'asymptote observée pour les adjectifs féminins singuliers (sec. 5.2.3.1).

S'il n'y a pas de corrélation entre la fréquence lexicale et les taux de défaut de <e> des lemmes, il convient en revanche de noter que les taux de ces défauts varient selon les lemmes, avec d'une part des lemmes qui n'apparaissent quasiment jamais avec leur <e> (« lors » : 99,9% de formes sans <e>; « mar » : 97% de formes sans <e>), des lemmes qui apparaissent majoritairement sans <e> (« onque » : 88% de formes du type *onc*; « or » : 77% de formes sans <e>; « desor » : 74% de formes sans <e>), un lemme au taux équilibré (« encore » : 57% de formes du type *encor*) et deux lemmes dont les formes majoritaires conservent leur <e> (« derriere » : 32% de formes du type *derrier*; « ariere » : 16% de formes du type *arier*).

7.2.3.4 Fréquence, collocations et défauts de <e>

Étant donnés les taux de défaut de <e> très élevés de 6 lemmes sur 9, il n'y a pas lieu de chercher une quelconque collocation qui favoriserait ces taux. Un examen manuel des occurrences en défaut de <e> de « ariere » et « derriere » ne nous a pas non plus permis de repérer l'effet d'une éventuelle collocation. En revanche, cet effet est sensible pour les occurrences *ensembl*, non seulement limitées à *Roland*, mais également circonscrites à la précession de la préposition *od*.

Ensembl od. Les formes du lemme « ensemble » de *Roland* se présentent sous la forme d'une distribution remarquable : les 17 formes en défaut de <e> sont placées devant *od* (ou *ot*), préposition issue du latin *APUD* (7.20 - 7.24).

- (7.20) *ensembl od lui rollant e oliver*
« avec lui sont Roland et Olivier »
(*Roland*, v. 104, trad. SHORT 1990)
- (7.21) *ensembl od els li quens rollant i vint*
« et avec eux le comte Roland s'en vint »
(*Roland*, v. 175, trad. *id.*)
- (7.22) *ses meillors humes enmeinet ensembl od sei*
« il y emmène ses principaux vassaux »
(*Roland*, v. 502, trad. *id.*)
- (7.23) *de ses parenz ensembl od lui tels trente*
« il y a trente de ses parents avec lui »
(*Roland*, v. 3781, trad. *id.*)
- (7.24) *de ses parenz ensembl od lui tels trente*
« et, avec lui, trente de ses parents »
(*Roland*, v. 1410, trad. *id.*)

En revanche, aucune des six formes sans défaut de <e> n'est en précession immédiate de *od* (7.25 - 7.30). Par ailleurs, parmi les six formes sans défaut

20: Dont l'une échappe aux contextes interrogés par *Comparalem*, l'initiale vocale étant <i>.

de <e>, trois sont placées devant initiale vocalique,²⁰ ce qui signifie que ce n'est pas l'initiale vocalique qui déclenche la forme défectueuse *ensembl*, mais bien la succession par la préposition *od* et nul autre qu'elle.

- (7.25) *ki que l cumpert venuz en sunt ensemble*
« Ils sont aux prises : c'est à qui le paiera. »
(*Roland*, v. 1635, trad. SHORT 1990)
- (7.26) *e sunt ensemble plus de cinquante milie*
« ils sont en tout plus de cinquante mil »
(*Roland*, v. 1919, trad. *id.*)
- (7.27) *ensemble avum estét e anz e dis*
« Des jours, des ans nous avons été ensemble »
(*Roland*, v. 2028, trad. *id.*)
- (7.28) *ensemble avruns e le ben e le mal*
« nous connaissons ensemble le meilleur et le pire »
(*Roland*, v. 2140, trad. *id.*)
- (7.29) *plus de cent milie s en adubent ensemble*
« plus de cent mille [Francs] s'arment tous ensemble »
(*Roland*, v. 3000, trad. *id.*)
- (7.30) *de ses parenz ensemble i out trente*
« il y a trente de ses parents avec lui »
(*Roland*, v. 3781, trad. *id.*)

Ainsi, les occurrences *ensembl* de *Roland* ne se trouvent que dans la collocation « ensemble od », et toutes les occurrences de la collocation « ensemble od » sont en défaut de <e> dans *Roland* (*ensembl od*), tandis qu'aucune autre occurrence, qu'elle soit devant initiale vocalique ou consonantique, ne génère de défaut de <e>. Cette collocation — ou du moins sa version qui a pour effet le défaut de <e> — ne fonctionne donc pas comme un *type*, c'est-à-dire comme une construction syntaxique dans laquelle les éléments ont encore une certaine liberté paradigmatique ou flexionnelle, mais semble rejoindre la catégorie des *tokens*, dont les composants sont strictement déterminés.

7.2.4 Distribution contextuelle

Les graphiques de séries temporelles en fig. 7.8.a-d (page suivante) représentent le taux de défaut de <e> pour chaque texte, pour les lemmes retenus à l'exception de *ensemble* et, comme pour le reste des données désormais, des formes avec '-s adverbial'. Les graphiques en (c) et (d) représentent le nombre d'occurrences des huit lemmes par la taille des points, et leurs mesures statistiques et régressions sont pondérées par ce nombre d'occurrences, ce qui n'est pas le cas en (a) et (b). Sur l'ensemble de ces graphiques, on observe la même tendance à la hausse du taux de défaut, mais la corrélation linéaire n'est jamais satisfaisante et significative.²¹ Surtout, comme nous le développerons *infra* (sec. 7.3.4), cette tendance à la hausse qui s'observe sur l'ensemble de nos données, n'est pas reproduite une fois les données anglo-normandes séparées des données continentales : la croissance de la régression linéaire est un effet diatopique et ne s'observe ni dans le sous-ensemble anglo-normand ni dans le sous-ensemble continental. Par ailleurs, dans chaque contexte, et que la pondération soit appliquée ou non, on retrouve une nouvelle

21: __#C	(a)	$\rho = 0,06$ $p = 0,44$
	(c)	$\rho = 0,11$ $p = 0,07$
__#V	(b)	$\rho = 0,09$ $p = 0,16$
	(d)	$\rho = 0,13$ $p = 0,03$

En (d), la mesure est significative ($p < 0,005$), mais le coefficient de corrélation est bien trop faible ($\rho = 0,13$.)

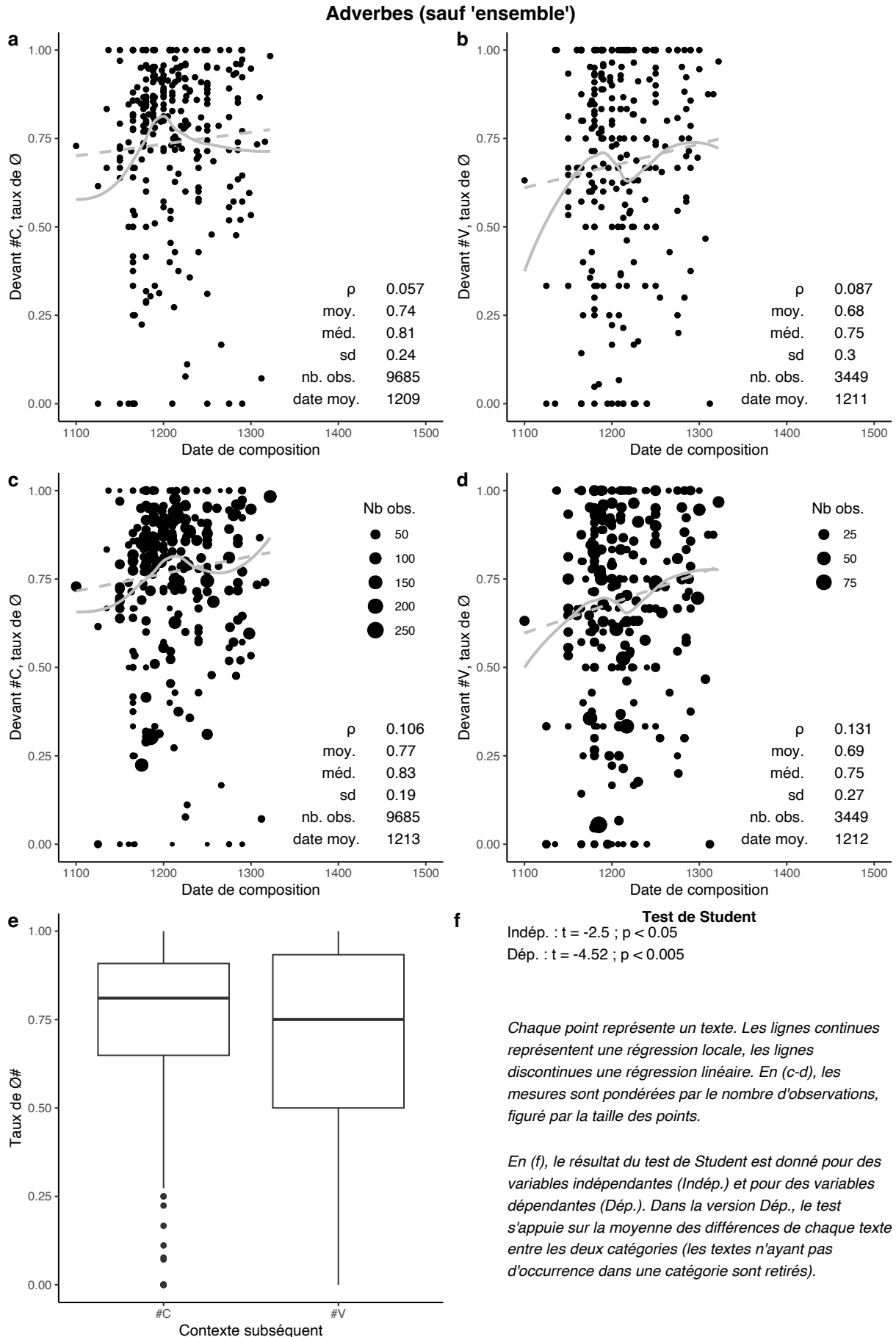


FIGURE 7.8 – Distribution contextuelle des variantes d'adverbes en défaut de <math>\langle e \rangle</math>.

fois un pic de la régression locale, mais son dessin est différent selon le contexte : devant initiale consonantique, la régression locale culmine en 1200 (a : sans pondération) ou peu après (c : avec pondération), tandis que devant initiale vocalique, la régression linéaire culmine deux fois, une première fois avant 1200 et une seconde à la fin du XIII^e s. Ce second pic est directement lié à la disparition presque totale de textes ayant 0% de formes en défaut de <e> après 1250.

En première observation, il convient de noter que les taux moyens de défaut de <e> sont supérieurs devant initiale consonantique, avec une moyenne de défaut de <e> de 74% devant initiale consonantique (moy. pond. = 77%), contre une moyenne de 68% devant initiale vocalique (moy. pond. = 69%). La distribution des données est également légèrement plus cohérente devant initiale consonantique, avec un écart-type de 0,24 devant initiale consonantique (σ pond. = 0,19), contre 0,3 devant initiale vocalique (σ pond. = 0,27). On s'attend donc à trouver, en statistique inférentielle, un effet du contexte. Comme indiqué en fig. 7.8.e-f cet effet, s'il est significatif ($p < 0,05$), n'est pas très intense ($t = -2,5$) : on ne trouve que légèrement plus de défauts de <e> devant initiale consonantique que devant initiale vocalique. Cet effet est plus marqué ($t = -4,52$) et encore plus significatif ($p < 0,005$) si l'on s'en tient aux seuls textes qui disposent d'occurrences dans les deux contextes, ce qui laisse à penser qu'une partie de la faiblesse de l'effet mesurée par le test de Student à variables indépendantes ($t = -2,5$) est due aux textes qui ne disposent que d'occurrences dans l'un ou l'autre des deux contextes, ce qui est probablement lié à un faible nombre d'occurrences de ces textes (cf. σ inférieur après application de la pondération). Au demeurant, dans les deux versions du test, la force de l'effet est inférieure à ce que nous avons noté pour les adjectifs féminins singuliers (Student Indép. $t = 4,84$; Student Dép. $t = 5,48$)²² et considérablement inférieure à ce que nous avons noté pour les prépositions (Student Indép. $t = 80,25$; Student Dép. $t = 79,54$).²³ La direction de l'effet est également inverse : alors que, pour toutes les catégories étudiées jusqu'ici, c'est le contexte prévoicalique qui rencontre le plus de défauts de <e> (relativement au nombre d'occurrences des lemmes retenus), ce qui se traduit par une valeur- t positive, pour les adverbes c'est l'inverse : c'est bien en contexte préconsonantique que l'on rencontre le plus de défauts de <e> (la valeur- t est négative).

Néanmoins, la valeur- p ne fait qu'indiquer la possibilité de rejeter l'hypothèse nulle (ici : le contexte n'a pas d'influence sur la distribution) au niveau de la statistique inférentielle, et ne dit rien de la pertinence réelle, dans l'analyse linguistique, des résultats. Étant donné le nombre et la distribution des données, la valeur- p indique ici qu'il est fortement (test à variable indépendantes) voire très fortement (test à variables dépendantes) improbable que l'hypothèse nulle soit vraie. Il y a donc, au niveau de la description des données, un effet probable du contexte subséquent, par la probable négation de l'hypothèse nulle d'absence d'effet du contexte. Néanmoins, au niveau de l'analyse linguistique, cet effet est à la fois très faible (plus faible que pour les adjectifs) et fortement improbable. En effet, alors que pour les autres parties du discours, le contexte prévoicalique a un effet positif sur les taux de défaut de <e>, c'est ici l'inverse. Non seulement il paraît hautement improbable que le contexte puisse inverser son effet pour une catégorie grammaticale

22: Cf. fig. 5.10, p. 154.

23: Cf. fig. 6.12, p. 206.

donnée, mais cette inversion est aussi phonologiquement improbable, dans la mesure où elle amènerait à favoriser la création de points de contact entre des consonnes à la frontière de mots, alors qu'un schwa est disponible à cet endroit pour éviter ce contact. L'hypothèse la plus probable, étant donné ce qui vient d'être dit, est que cet effet statistique faible et inverse trahit l'absence d'effet phonologique réel.

Enfin, si nous avons établi en sec. 7.2.3.3 que le taux de défaut était fortement variable selon les lemmes, il convient de noter que l'absence d'effet significatif du contexte vaut non seulement pour tous les lemmes pris ensemble, mais est aussi valable pour la moitié des lemmes retenus. La table 7.6 reproduit les données lemme par lemme et contexte par contexte. Elle permet d'observer que les taux de défaut de <e> sont égaux ou équivalents pour la moitié des lemmes.

TABLE 7.6 – Relation entre fréquence et taux de défaut de <e> par contexte, pour chaque lemme de préposition.

lemma	__#V			__#C			__#{V, C}	
	nb.nonE	nb.E	Tx_zero	nb.nonE	nb.E	Tx_zero	nb	Tx_zero
or ^a	1347	474	74%	3305	901	79%	6027	77%
lors	592	2	99,7%	2275	0	100%	2869	99,9%
encore	264	220	55%	702	505	58%	1691	57%
arriere	46	271	15%	105	514	17%	936	16%
onque	38	3	93%	449	62	88%	552	88%
desor	19	42	31%	382	97	80%	540	74%
mar	66	0	100%	188	9	95%	263	97%
derriere ^b	15	50	23%	66	125	35%	256	32%

a. or_relemma

b. derriere_relemma

En revanche, pour les quatre autres lemmes, on peut observer une différence de taux de défaut en fonction du contexte. Pour « onque » et « mar », on trouve légèrement plus de formes défectueuses devant initiale vocalique (« onque » : 93% ; « mar » : 100%) que devant initiale consonantique (« onque » : 88% ; « mar » : 95%). De fait, « mar » n'a aucune forme non défectueuse devant initiale vocalique, mais n'a que 9 formes non défectueuses devant initiale consonantique, sur 188 formes. Non seulement la différence quantitative est faible, mais il n'est pas sûr que ce résultat soit véritablement significatif sur le plan linguistique (au sens d'une contrainte contre <e> devant initiale vocalique), étant donné le faible nombre de formes non défectueuses du lemme. Pour « onque », en revanche, 3 formes non défectueuses devant initiale vocalique (sur 38 formes) s'opposent à 62 formes non défectueuses devant initiale consonantique (sur 449 formes). Si ce résultat semble mieux consolidé, il ne témoigne ni d'une contrainte absolue contre <e> en contexte prévoicalique ni d'un effet aussi fort que celui noté pour les prépositions, et dans les deux contextes le taux de défaut reste très élevé.

Pour « derriere » et « desor », c'est l'inverse : c'est le contexte préconsonantique qui est associé aux plus hauts taux de défaut de <e> (« derriere » : 35% __#C vs 23% __#V ; « desor » : 80% __#C vs 31% __#V). Si la différence est patente pour « desor », il faut néanmoins rappeler que l'effet du contexte est très faible et non significatif pour ces mêmes lemmes en fonction de préposition (PRÉP « sor » et « desor » : $t = 0,02$; $p = 0,98$; cf. fig. 6.19), et qu'il est difficile d'accorder une interprétation linguistique

à une distribution favorisant les défauts devant initiale consonantique. Au demeurant, ces quatre lemmes qui semblent montrer une différence entre les contextes prévocaliques et préconsonantiques sont les moins fréquents de nos huit lemmes, ne dépassant pas 552 occurrences pour « onque », contre entre 936 et 6027 occurrences pour les quatre autres lemmes. Si cela peut éventuellement témoigner d'un effet fréquentiel sur les taux contextuels (et non sur les taux non contextuels, cf. sec. 7.2.3.3), cela peut aussi témoigner d'une simple insuffisance de données pour ces lemmes. Enfin, s'il fallait y voir un effet de la fréquence lexicale (seuls les lemmes à faible fréquence conservent une distribution contextuelle hétérogène), il faudrait néanmoins convenir que cette conservation de l'effet contextuel fonctionne aussi bien en faveur du contexte prévocalique que préconsonantique, ce qui interdit d'en tirer une conclusion phonologique en tant que telle. Ainsi, dans la mesure où certains lemmes favorisent le défaut dans l'un contexte et d'autres dans l'autre contexte, et où la moitié des lemmes favorisent voire généralisent le défaut dans les deux contextes, ce qui semble à l'œuvre ici, plutôt qu'une contrainte phonologique, semble être le hasard.

7.2.4.1 En somme

L'effet du contexte subséquent sur les adverbes est significatif, mais faible, et c'est le contexte préconsonantique qui favorise le plus les défauts de <e>. Cette significativité statistique indiquée par la valeur-*p* n'implique pas une significativité linguistique ; la faiblesse de la valeur-*t* et l'improbabilité que le contexte puisse inverser son effet sur une partie du discours et que le contexte préconsonantique puisse favoriser les défauts de <e> nous amènent à la conclusion que le contexte n'a pas d'effet linguistique patent sur les défauts de <e> des adverbes. Par ailleurs, si 2 lemmes semblent préférer le défaut devant initiale vocalique, deux lemmes le préfèrent devant initiale consonantique, et quatre lemmes ne font aucune différence ; ce qui augmente d'autant l'improbabilité d'un effet linguistique : le contexte devrait non seulement inverser son effet sur une partie du discours, mais également de manière différenciée en fonction des lemmes au sein de cette partie du discours. Les 4 lemmes favorisant le défaut dans l'un des deux contextes étant ceux dont la fréquence est la plus faible, cela ne semble être dû qu'à la dimension stochastique du phénomène à l'œuvre.

7.2.5 Distribution diatopique

Les cartes en fig. 7.9 et 7.11 représentent le taux moyen de défaut de <e> par texte pour chaque région, en synchronie, respectivement devant initiale consonantique et devant initiale vocalique ; comme le reste de nos données, les formes en '-s adverbial' sont désormais omises. Devant initiale consonantique (7.9), les taux sont sensiblement plus élevés dans une petite moitié Sud-Est du domaine : en-deçà d'une ligne partant de la région Indre, Cher et allant jusqu'à la Wallonie, on ne trouve aucun taux inférieur à la moyenne des taux régionaux (78%). D'autres taux supérieurs sont relevés au-dessus de cette ligne, principalement au Nord du domaine (Nord : 93%, Hainaut : 86%), en Normandie (78%) et en région Vendée, Deux-Sèvres (75%). En revanche, le long d'un arc de

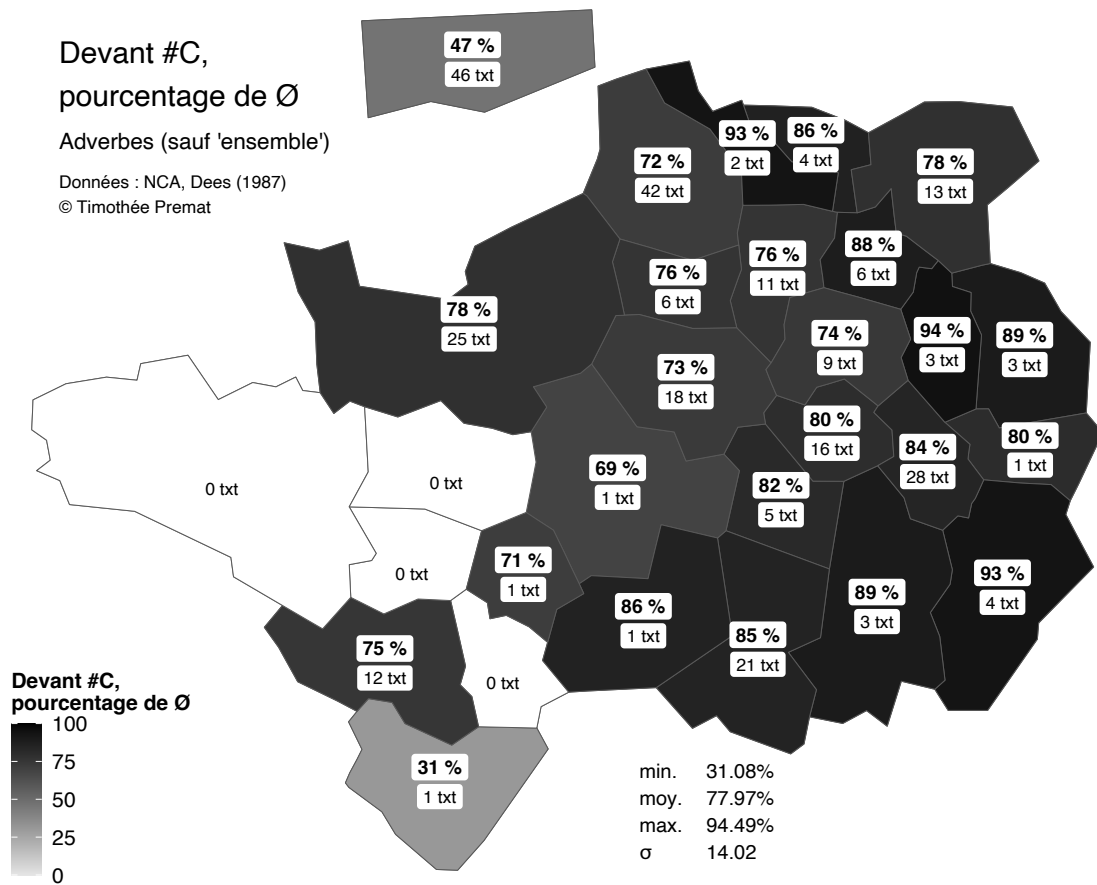


FIGURE 7.9 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, adverbes (sauf « ensemble ») suivis d'une initiale consonnante.

cercle partant de la Charente-Maritime et aboutissant à l'Angleterre, en passant par le centre du domaine, les taux sont tous inférieurs à la moyenne, et les deux taux les plus bas se situent aux extrémités de cet arc (Charente-Maritime : 31%, Angleterre : 47%). Le nombre de textes attestant nos lemmes ne semble pas avoir une influence décisive, avec des régions bien dotées en textes dans les valeurs faibles (Angleterre : 46 textes), dans les valeurs élevées (région Nièvre, Allier : 21 textes) comme dans les valeurs moyennes (Normandie : 25 textes, Wallonie : 13 textes).

La carte en fig. 7.10, qui reproduit la fig. 7.9 en faisant abstraction des régions disposant de moins de cinq textes, ne change pas fondamentalement la lecture de la distribution diatopique. Elle ne permet certes plus d'attester de la continuité de l'arc de cercle des taux inférieurs à la moyenne, mais elle confirme le pôle inférieur à la moyenne entre l'Angleterre et la Marne et la région parisienne. Parallèlement, si cinq des six régions les plus occidentales disposent de moins de cinq textes et disparaissent donc de cette carte, les taux sont cohérents avec ceux des régions qui sont situées à leur gauche; si l'on peut douter de la représentativité réelle de leurs taux les plus élevés, il n'y a donc pas de raison

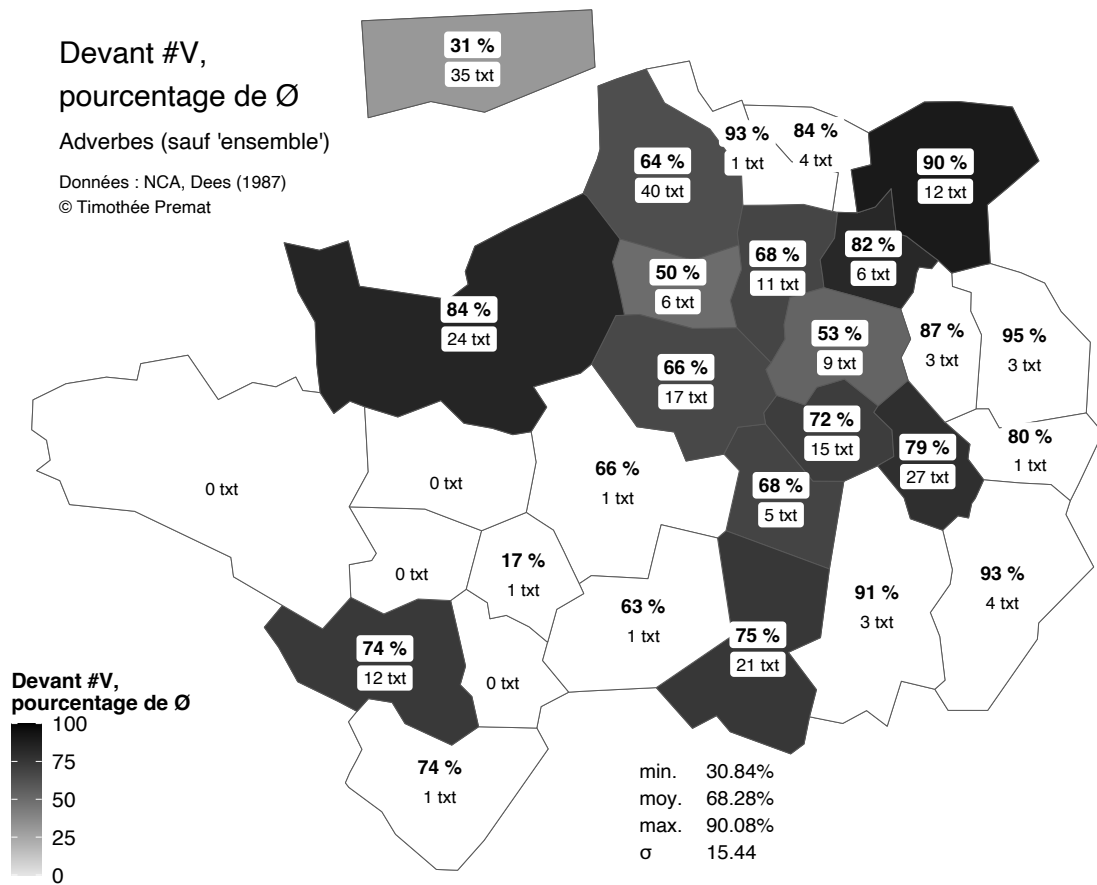


FIGURE 7.12 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, adverbes (sauf ensemble) suivis d'une initiale vocalique, et à l'omission des régions disposant de moins de cinq textes.

La carte en fig. 7.12 qui reproduit la fig. 7.11 en faisant abstraction des régions disposant de moins de cinq textes, témoigne de la même distribution que devant initiale consonantique. Si les taux extrême-orientaux sont peu consolidés au niveau du nombre des textes, ils sont bordés à l'Ouest par des régions dont les taux sont supérieurs ou égaux à la moyenne (sauf dans la Marne : 53%), aussi il n'y a pas de raison objective de remettre en question le fait que l'Est du domaine dispose de taux supérieurs au centre du domaine.

Si l'on cumule les deux contextes, la distribution diatopique prend la forme d'un axe regroupant les taux inférieurs à la moyenne, allant de l'Angleterre jusqu'à la Charente-Maritime (__#C) ou jusqu'aux régions Indre, Cher et Indre-et-Loire (__#V), en passant par le centre du domaine. Cet arc de cercle est enserré entre deux autres axes, où les taux sont toujours supérieurs à la moyenne. L'axe oriental, partant du Nord et allant jusqu'à la région Indre, Cher (__#C) ou la région Nièvre, Allier (__#V), dispose des taux les plus élevés, concentrés dans son tiers septentrional et méridional. L'axe occidental, discontinu par l'absence de données, dispose de taux plus proches de la moyenne, et se réduit à deux régions,

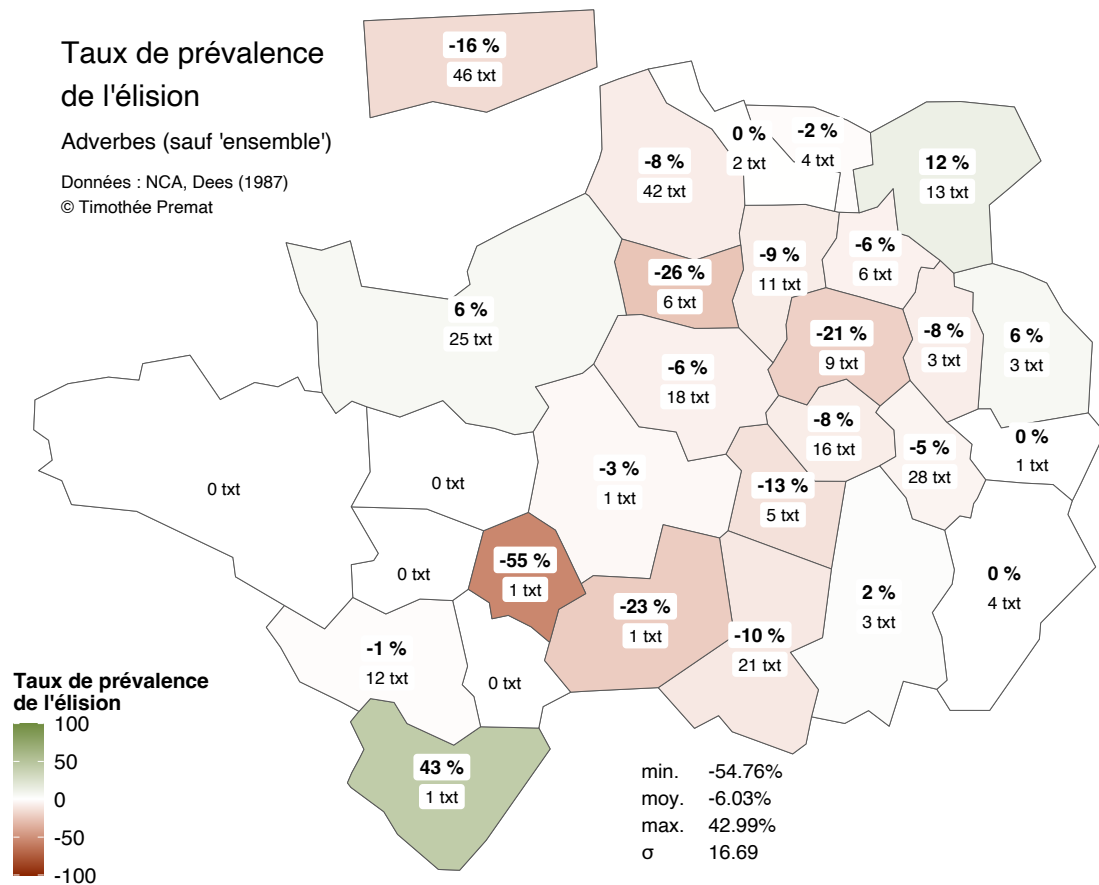


FIGURE 7.13 – Distribution diatopique de la prévalence de l'élision, sur l'ensemble de la période considérée, adverbes (sauf « ensemble ») suivis d'une initiale vocalique.

relativement bien dotées par ailleurs en nombre de textes (Normandie : 25 textes __#C, 24 textes __#V ; région Vendée, Deux-Sèvres : 12 textes).

La carte en fig. 7.13 représente le degré de prévalence de l'élision, c'est-à-dire le taux de défaut devant initiale vocalique moins le taux de défaut devant initiale consonnantique pour chaque région. Si le défaut de <e> est plus prononcé devant initiale consonnantique que devant initiale vocalique dans la plupart des régions (moy. = -6%), il apparaît une certaine hétérogénéité ($\sigma = 17$), avec une distribution remarquable, où trois des quatre 'coins' du domaine continental sont les seules régions à favoriser le défaut de <e> devant initiale vocalique plutôt que devant initiale consonnantique. Cette distribution concerne la Charente-Maritime (43%, 1 texte), la Normandie (6%, 25 textes), ainsi que la Wallonie (12%, 13 textes) et la région Moselle, Meurthe-et-Moselle (6%, 3 textes). Si deux de ces quatre régions disposent de peu de textes, la Normandie et la Wallonie disposent d'un nombre de textes satisfaisant. Ainsi, le faible effet du contexte préconsonnantique comme favorisant le défaut de <e> que nous avons noté en sec. 7.2.4 *supra* n'est pas stable dans l'espace : non seulement l'effet semble négligeable ou nul dans certaines régions

(Sud-Est : Bourgogne : 2% ; Franche-Comté 0% et Vosges : 0% ; Nord : 0%, Haunaut : -2%), mais il résulte du nivellement entre des régions où l'effet est plus ou moins prononcé au centre du domaine et en Angleterre (de -3% dans l'Orléanais, 1 texte et -5% dans la Haute-Marne, 28 textes, à -55% dans l'Indre-et-Loire, 1 texte et -26% dans l'Oise, 6 textes) avec des régions où c'est l'effet inverse qui est observé (Charente-Maritime, Normandie, Wallonie, région Moselle, Meurthe-et-Moselle). Cet effet, dont nous avons déjà écarté la significativité linguistique, est donc par ailleurs très disparate dans l'espace.

7.2.5.1 En somme

En synchronie, la distribution diatopique des taux régionaux de défaut de <e> est réductible à deux caractéristiques principales. D'une part, on note l'existence d'un arc de cercle partant du Sud-Ouest et se terminant en Angleterre, en passant par le centre du domaine, qui regroupe des taux inférieurs à la moyenne ; de part et d'autre de cet arc, les taux sont toujours supérieurs à la moyenne. D'autre part, cet arc de cercle est plus large devant initiale vocalique qu'il ne l'est devant initiale consonantique, mais il descend moins loin au Sud-Ouest devant initiale vocalique qu'il ne le fait devant initiale consonantique. La comparaison des taux de défaut devant initiale vocalique et devant initiale consonantique montre un territoire central et anglo-normand où le défaut devant initiale consonantique domine plus ou moins franchement le défaut devant initiale vocalique, tandis que quelques régions, singulièrement périphériques, témoignent d'une domination du défaut de <e> devant initiale vocalique.

7.2.6 Interface des distributions diatopiques et diachroniques

Les cartes en fig. 7.14, 7.16, 7.18 et 7.20, affichées dans les pages suivantes, présentent la moyenne régionale des taux de défaut de <e> des textes de chaque région, par période d'un demi siècle et de 72 ans pour la dernière période. De manière complémentaire, les cartes fig. 7.15, 7.17, 7.19 et 7.21 présentent, pour ces mêmes périodes, l'indice de prévalence de l'élision, c'est-à-dire la soustraction du taux moyen régional de défaut devant initiale consonantique au taux moyen régional de défaut devant initiale vocalique ; les indices négatifs, nombreux pour cette catégorie, signifient que l'on trouve plus de défauts de <e> devant initiale consonantique que devant initiale vocalique. Nous convenons, ici, que l'observation croisée des cartes de prévalence de l'élision et des cartes de taux de défaut devant initiale vocalique n'est pas des plus pratiques pour rétablir le taux de défaut devant initiale consonantique. Les sorties graphiques de Comparalem ont été implémentées pour des données où l'élision prévaut, ce qui n'est pas le cas ici. Pour chaque période, nous commençons donc par décrire l'évolution diachronique de la distribution diatopique des défauts de <e> devant initiale consonantique, puis la comparons, par le biais des cartes de prévalence de l'élision, à la situation devant initiale consonantique.

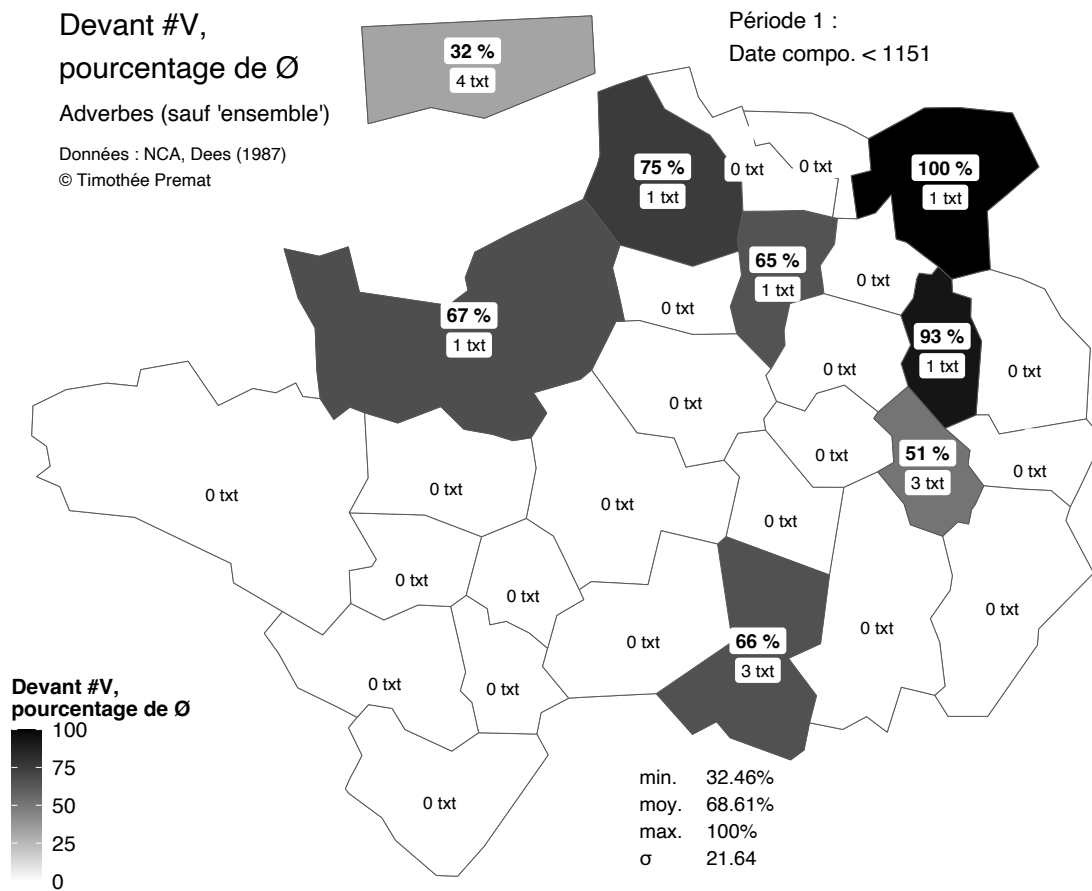


FIGURE 7.14 – Distribution diatopique des alternances <e>/Ø, période 1 (1100-1150), adverbes (sauf « ensemble ») suivis d'une initiale vocalique.

7.2.6.1 1100-1150

__#V. Devant initiale vocalique et dans la première moitié du XII^e s. (fig. 7.14), le défaut de <e> est attesté partout où nos lemmes sont attestés. La faiblesse caractéristique de l'Angleterre est déjà observable, quoiqu'elle soit surévaluée par l'omission des occurrences *ensembl* de *Roland*. Le domaine continental reste assez contrasté, avec des taux allant de 51% en Haute-Marne (3 textes) à 100% en Wallonie (1 texte). On ne peut évidemment voir se dessiner entièrement la distribution synchronique évoquée *supra*, étant donné la rareté des textes et, surtout, des régions. On note néanmoins que le Nord-Est (Wallonie, région Moselle, Meurthe-et-Moselle), se distingue par des taux maximaux (respectivement 100% et 93%) et, même si ces régions ne comportent qu'un seul texte chacune, elles sont aussi des régions ayant des taux élevés en synchronie (90% et 95% en fig. 7.11 *supra*).

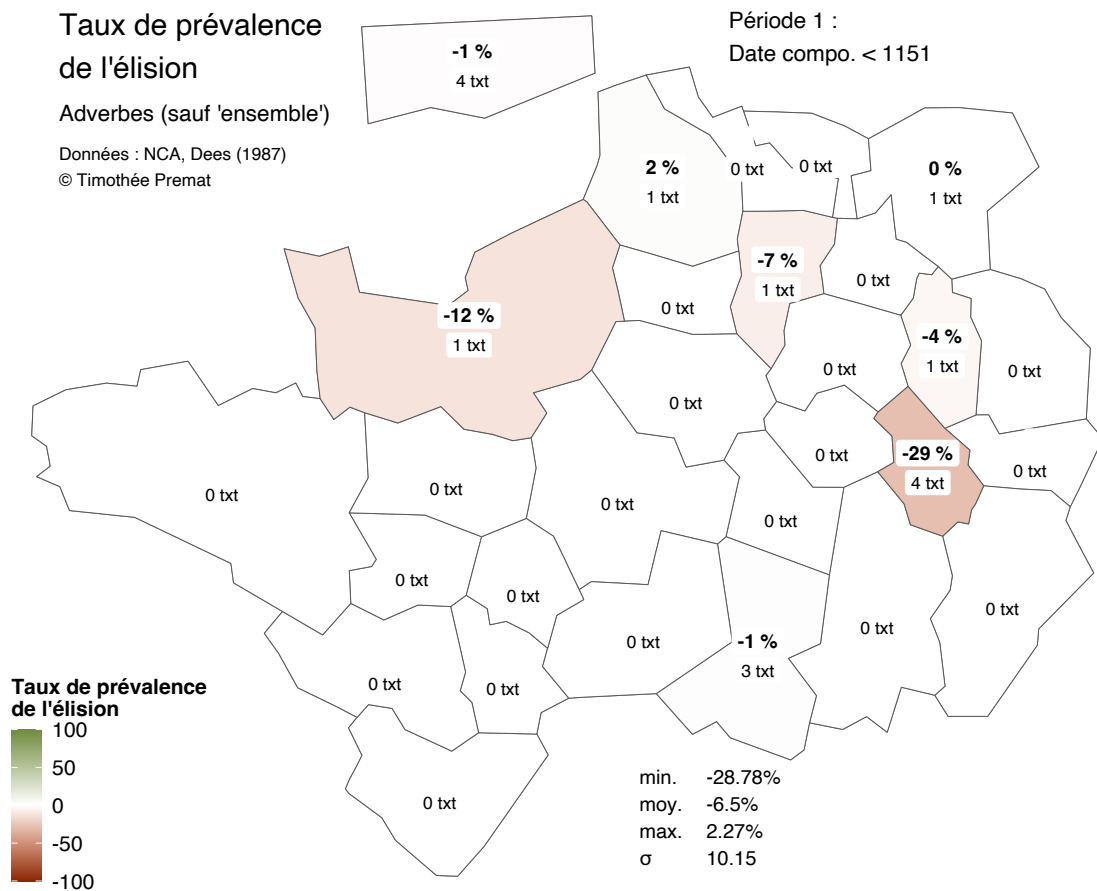


FIGURE 7.15 – Distribution diatopique de la prévalence de l'élision, période 1 (1100-1150), adverbes (sauf « ensemble ») suivis d'une initiale vocalique.

Prévalence de l'élision et __#C. Dans la première moitié du XII^e s., aucune région ne montre des taux de défaut de <e> devant initiale vocalique significativement plus élevés que devant initiale consonantique (7.15). Seule la région Somme, Pas-de-Calais dispose d'un taux de défaut de <e> devant initiale vocalique (75%) légèrement plus élevé que devant initiale consonantique (73%). Au contraire, la Normandie et la Haute-Marne montrent des taux de défaut de <e> significativement plus élevés devant initiale consonantique que devant initiale vocalique (Normandie : 79% __#C vs 67% __#V ; Haute-Marne : 71% __#C vs 51% __#V). Aussi, si, devant initiale vocalique, ces deux régions ne permettent pas véritablement d'observer déjà la distribution observée en synchronie, elles se rapprochent déjà plus de cette distribution synchronique devant initiale consonantique, en se rattachant clairement aux axes orientaux et occidentaux qui favorisent les défauts de <e>.

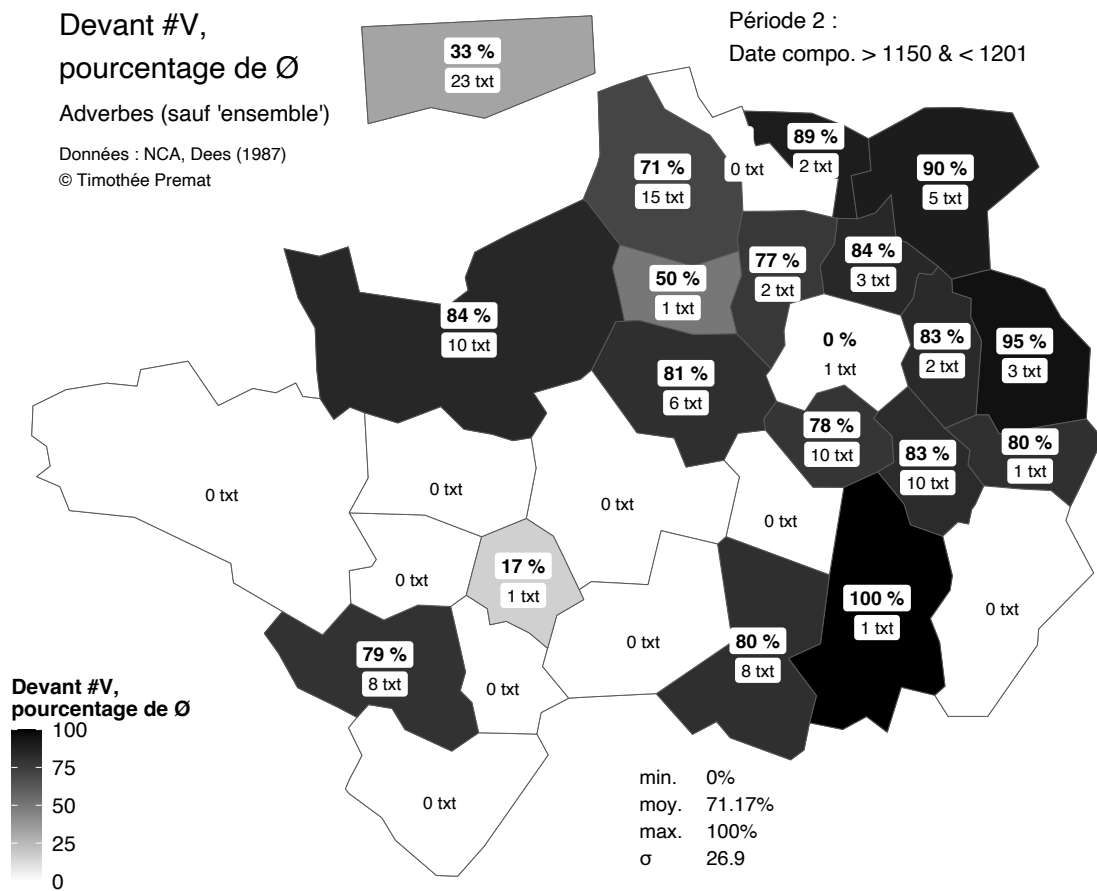


FIGURE 7.16 – Distribution diatopique des alternances <e>/Ø, période 2 (1151-1200), adverbes (sauf « ensemble ») suivis d'une initiale vocalique.

7.2.6.2 1151-1200

#V. Dans la seconde moitié du XII^e s. et devant initiale vocalique (fig. 7.16), malgré un certain mitage des données, la situation synchronique commence à être repérable, avec des taux de défaut de <e> plus élevés à l'Est et au Nord-Est du domaine qu'au centre du domaine. Le nombre de textes comportant nos lemmes joue ici clairement un rôle de filtre, notamment dans la Marne, dont le taux de 0% ne provient que d'un seul texte. L'Angleterre demeure également marginale, avec le second taux le plus bas (33%, 23 textes), derrière la région Indre-et-Loire (17%, 1 texte), dont le taux est à la fois trop faible et trop peu consolidé par le nombre de textes pour être vraiment pertinent. Néanmoins, dans le centre du domaine, plusieurs régions continuent d'avoir des taux substantiellement supérieurs à la moyenne des taux régionaux (71%), avec notamment la région parisienne (81%, 6 textes) et l'Aisne (77%, 2 textes). Si l'on ne rejette pas les taux de ces régions sur le fondement de leur faible nombre de textes, il faut remarquer qu'en synchronie elles appartiennent au centre de l'arc de cercle des taux inférieurs à la moyenne (avec respectivement 66, et 68% de défauts devant #V); dans la seconde moitié du XII^e s.,

textes, etc.). De fait, alors que la partie centrale de l'arc de cercle aux taux inférieurs à la moyenne ne se dessine que partiellement, avec une intensité moindre et un mitage important devant initiale vocalique, les indices de prévalence de l'élision négatifs dans cette région indiquent que ce n'est pas le cas devant initiale consonantique, où ces régions se démarquent par des taux moins proches de la moyenne.

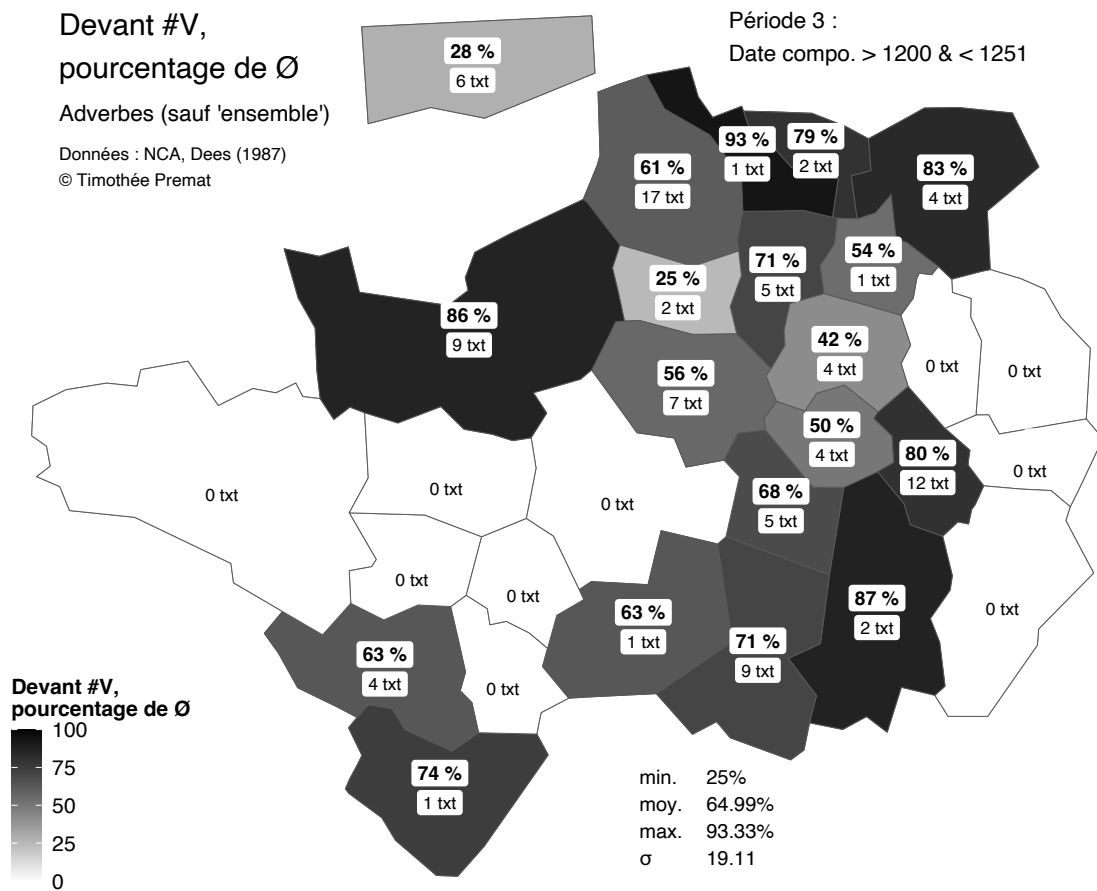


FIGURE 7.18 – Distribution diatopique des alternances <e>/Ø, période 3 (1201-1250), adverbes (sauf « ensemble ») suivis d’une initiale vocalique.

7.2.6.3 1201-1250

__#V. Dans la première moitié du ^{xiii}e s., la situation se rapproche de celle observée en synchronie : l’arc de cercle Angleterre - centre - Sud-Sud-Ouest du domaine se dessine, avec des taux inférieurs à la moyenne (fig. 7.18). Cet arc est néanmoins discontinu du fait d’une absence de données dans l’Orléanais, l’Indre-et-Loire et la Vienne (la Vienne n’a aucun texte dans le corpus). Si les régions les plus orientales sont absentes, elles sont bordées par des régions dont les taux sont supérieurs à la moyenne (moy. = 65% ; Bourgogne : 87%, Haute-Marne : 80%, Wallonie : 83%) et qui appartiennent donc à l’axe occidental aux taux de défaut de <e> élevés, soit par des régions dont les taux sont inférieurs à la moyenne, et qui appartiennent alors à la marge de l’arc de cercle décrit *supra* (Ardennes : 54%, Marne : 42%).

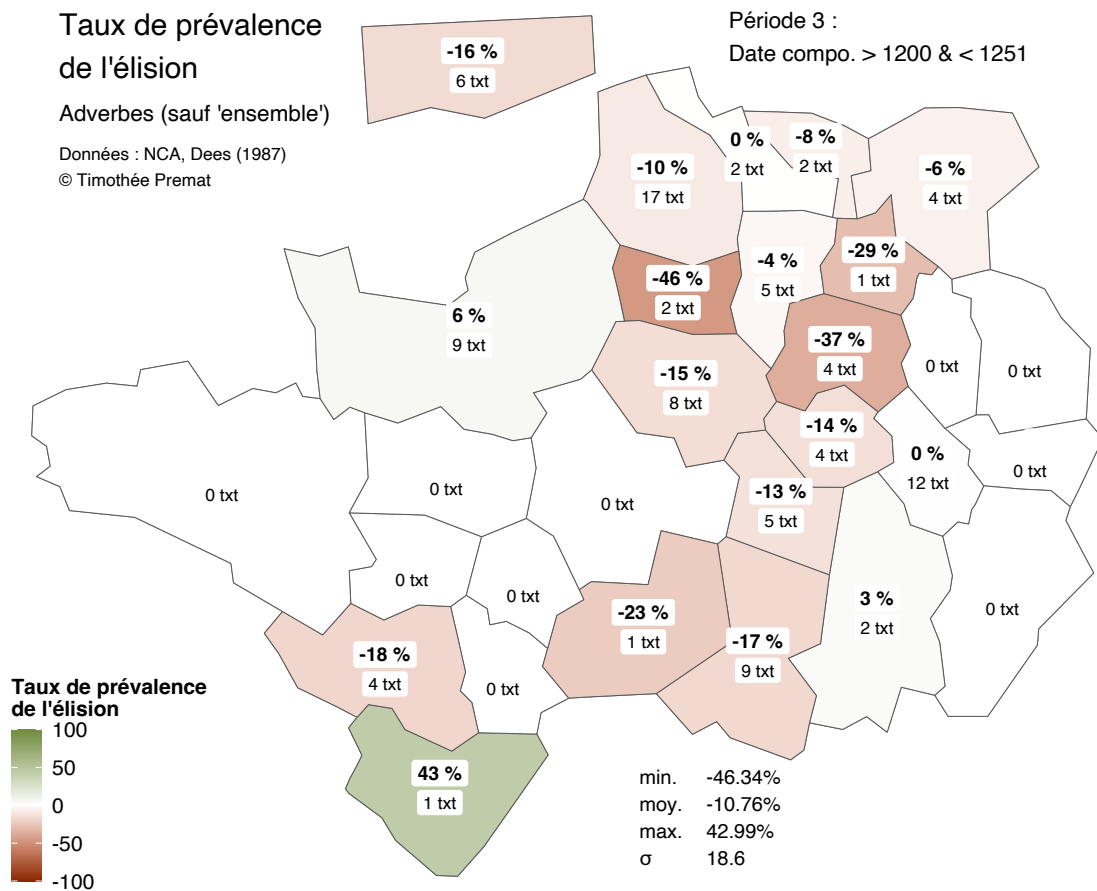


FIGURE 7.19 – Distribution diatopique de la prévalence de l'élision, période 3 (1201-1250), adverbes (sauf « ensemble ») suivis d'une initiale vocalique.

Prévalence de l'élision et __#C. La distribution du rapport entre taux de défaut devant initiale consonnantique et devant initiale vocalique se fait plus homogène dans cette première moitié du ^{xiii}e s. (fig. 7.19) que dans la période précédente : toutes les régions sauf deux favorisent les défauts de <e> devant initiale consonnantique plutôt que devant initiale vocalique, les exceptions étant la Charente-Maritime (31% __#C vs 74% __#V, 1 texte) et la Normandie (80% __#C vs 86% __#V, 9 textes). Les régions où l'effet favorable au défaut du contexte préconsonnantique est le plus marqué sont situés au Centre-Nord du domaine, avec l'Oise (71% __#C vs 25% __#V), la Marne (79% __#C vs 42% __#V) et les Ardennes (83% __#C vs 54% __#V). Aussi, il convient de considérer que, comme dans la seconde moitié du ^{xiii}e s., c'est le contexte préconsonnantique qui est le plus cohérent et qui représente déjà la situation la plus proche de la répartition synchronique observée en fig. 7.9 *supra*.

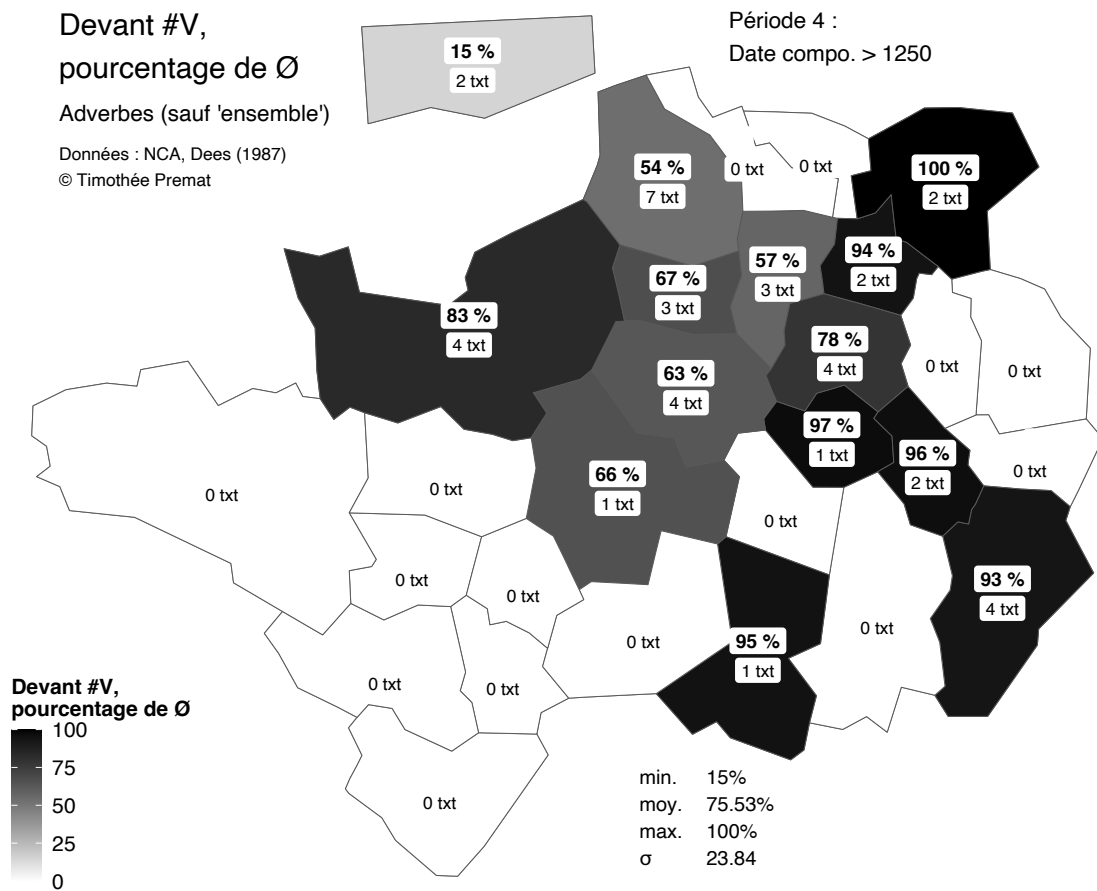


FIGURE 7.20 – Distribution diatopique des alternances <e>/Ø, période 4 (1251-1322), adverbes (sauf « ensemble ») suivis d’une initiale vocalique.

7.2.6.4 1251-1322

#V. Enfin, dans la seconde moitié du XIII^e s. et au début du XIV^e s., la situation devant initiale vocalique témoigne d’une absence de données dans de nombreuses régions (fig. 7.20). Néanmoins, dans les régions où nos lemmes sont attestés, les taux sont pleinement conformes avec ce que nous avons observé sur la carte synchronique en fig. 7.11 *supra* : l’Est et l’Ouest (ici réduit à la Normandie) disposent de taux supérieurs à la moyenne (moy. = 76%), avec la Marne à 78% de défaut de <e>, la Normandie à 83% et 6 régions dont le taux est supérieur ou égal à 94% (région Nièvre, Allier, Franche-Comté, Haute-Marne, Aube, Ardennes, Wallonie). Les Ardennes ont changé de statut : avec un seul texte à 51% de défaut en période 3, elles ont désormais deux textes et 94% de défauts, rejoignant ainsi la partie occidentale du domaine. Quant aux régions de l’arc de cercle des faibles taux, elles se divisent entre un taux anglo-normand extrêmement faible (15%) et des taux continentaux relativement homogènes, allant de 54 et 57% dans la région Somme, Pas-de-Calais et l’Aisne à 66 et 67% dans l’Orléanais et en région parisienne.

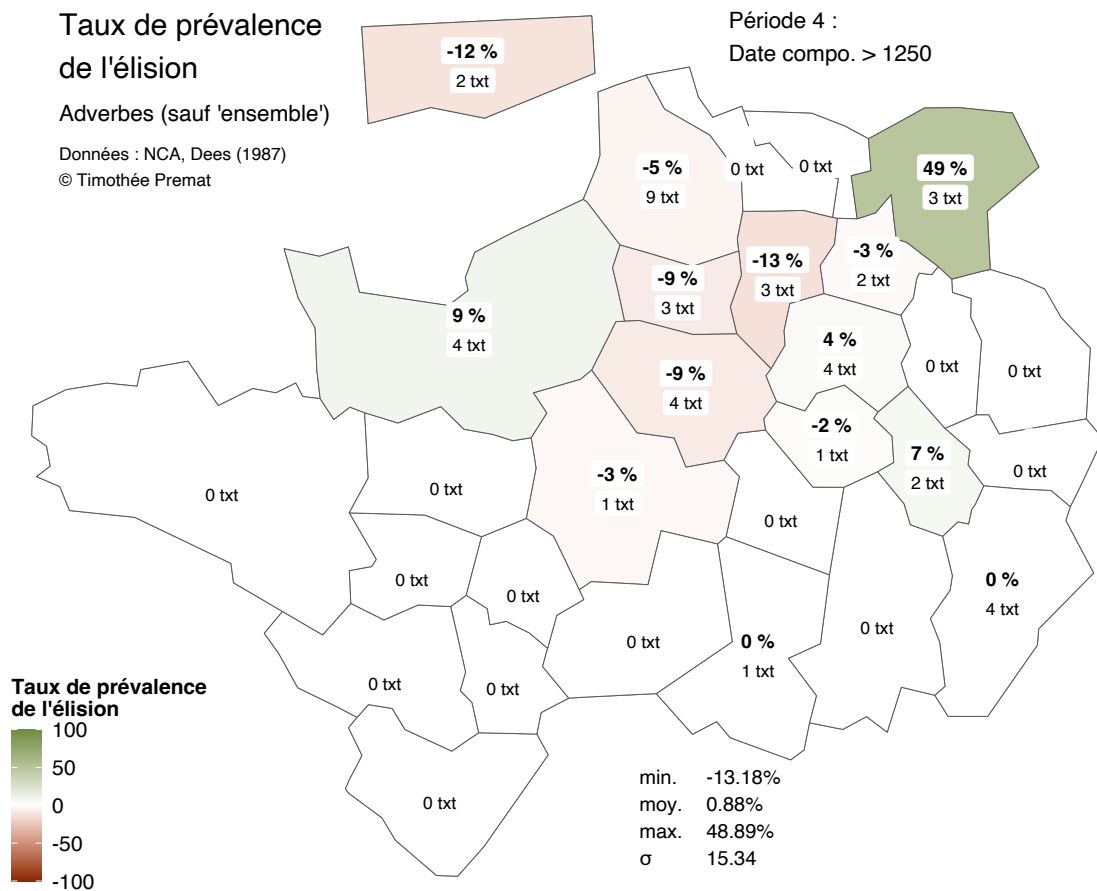


FIGURE 7.21 – Distribution diatopique de la prévalence de l'élision, période 4 (1251-1322), adverbes (sauf « ensemble ») suivis d'une initiale vocalique.

Prévalence de l'élision et $_ \#C$. Dans cette dernière période, hormis une valeur extrême en Wallonie, l'indice de prévalence de l'élision est relativement modéré (fig. 7.21), de -13% dans l'Aisne (70% $_ \#C$ vs 57% $_ \#V$) et -12% en Angleterre (27% $_ \#C$ vs 15% $_ \#V$) à 7% en Haute-Marne (89% $_ \#C$ vs 96% $_ \#V$) et 9% en Normandie (74% $_ \#C$ vs 83% $_ \#V$). Aussi, même là où c'est le contexte prévoicalique qui génère, relativement au nombre d'occurrences de chaque contexte, le plus d'occurrences en défaut de <e>, les valeurs devant initiale consonantique demeurent élevées. À l'inverse, deux des trois régions dont les taux de défaut étaient les plus bas devant initiale vocalique sont celles où l'indice de prévalence de l'élision est le plus négatif, ce qui signifie que leurs taux de défaut devant initiale consonantique, tout en étant toujours bas dans le cas de l'Angleterre (27% $_ \#C$), demeurent relativement proches de la moyenne. Hormis la situation exceptionnelle de la Wallonie, où le taux de défaut de <e> est presque deux fois plus faible devant initiale consonantique qu'il ne l'est devant initiale vocalique (51% $_ \#C$ vs 100% $_ \#V$), la situation générale est donc relativement conforme à ce qui a été observé en synchronie, malgré un certain mitage qui vient atténuer le contraste entre l'arc de cercle et les zones qui l'entourent, notamment

dû à des indices de prévalence de l'élision positifs en Normandie, en Haute-Marne et, dans une moindre mesure, dans la Marne.

7.2.6.5 En somme

La comparaison des différentes cartes de cette section permet de mettre en avant plusieurs éléments. D'une part, à toutes les périodes, le contexte préconsonantique semble plus cohérent (au niveau de sa distribution diatopique) et plus en adéquation avec la distribution synchronique que ne l'est le contexte prévocalique. La domination d'indices de prévalence de l'élision négatifs, parfois avec des taux très marqués, indique également que, diachroniquement, dans la plupart des régions, le contexte préconsonantique est en avance sur le contexte prévocalique pour ce qui est des taux de défaut de <e>.

Enfin, à différentes périodes, les marges et parfois quelques régions non marginales font état d'indices de prévalence de l'élision à valeur positive, ce qui indique que l'effet du contexte est fortement variable. Si, dans certains cas, ces indices positifs sont liés à de faibles nombres de textes, ce n'est pas toujours le cas.

Tous ces éléments amènent à renforcer le refus de significativité linguistique de l'effet du contexte en synchronie, étant donné que cet effet est extrêmement variable dans le temps et procède du nivellement d'une situation fort mouvante. En revanche, il convient de noter que le contexte préconsonantique semble devancer le contexte prévocalique ; si cet élément était, en quelque sorte, déjà visible sur les graphiques de séries temporelles en fig. 7.8 *supra*, il est marquant de constater que c'est aussi le cas en diatopie et que, mis à part quelques cas de mitage des données et, évidemment, de données manquantes à telle ou telle période, cette avance du contexte préconsonantique est globalement valable non seulement en diachronie mais également à l'interface entre diachronie et diatopie.

7.3 Des observations aux généralisations

Dans cette section, nous réduisons la profusion de données exposées *supra* à un ensemble de caractéristiques pertinentes. Nous commençons par reprendre ce que nous avons pu avancer à propos du rôle de la fréquence des lemmes, dans l'absolu et dans les constructions, sur les défauts de <e> (sec. 7.3.1), avant de revenir sur le rôle du contexte phonologique (sec. 7.3.2). En abordant la distribution diachronique des défauts (sec. 7.3.3), nous introduisons un facteur diatopique, ensuite repris dans l'étude de la diatopie en elle-même (sec. 7.3.4) et dans l'étude de la spécificité de l'anglo-normand (sec. 7.3.5). Enfin, nous interrogeons la place de la variation au sein de l'ensemble de la catégorie des adverbes, et avançons quelques comparaisons avec les autres catégories étudiées (sec. 7.3.6).

7.3.1 Lemmes atteints par le défaut de <e>

Comme indiqué en sec. 7.2.3, nos neuf lemmes sont attestés en défaut de <e> soit dès *Roland* soit dans la première moitié du XII^e s. Si nous n'avons retenu que neuf lemmes et, en définitive, huit après avoir remarqué que le lemme « ensemble » n'était défectueux que dans *Roland*, nous avons estimé en sec. 7.2.1 que ces huit lemmes représentaient approximativement la moitié des lemmes réunissant les conditions étymologiques, phonologiques et quantitatives nécessaires pour apparaître en défaut de <e>. Le faible nombre de nos huit ou neuf lemmes (un peu plus de 1% des lemmes adverbiaux) n'est donc pas faible parce que le défaut s'applique faiblement, mais parce que peu d'items sont susceptibles de l'exprimer. Ces observations peuvent être synthétisées par la généralisation en (7.31).

(7.31) **Adv : sélectivité lexicale.**

Nous estimons qu'à peu près la moitié des lemmes adverbiaux réunissant les conditions pour les défauts de <e> ont des occurrences en défaut de <e>.

Du fait que nos lemmes sont attestés très tôt dans le corpus, dès *Roland* pour plus de la moitié d'entre eux, et dans la première moitié du XII^e s. pour les autres, la variation étudiée semble être présente d'emblée, et non pas se développer, en termes de lemmes atteints, avec le temps. Il convient néanmoins de noter que les deux lemmes atteints le plus tardivement (quoi que précocement comparé aux autres catégories) par les défauts de <e> sont aussi ceux qui ont les taux les plus bas ; le lemme « derriere » n'apparaît en défaut de <e> qu'autour de 1137, tandis que le lemme « ariere » n'apparaît en défaut de <e> qu'autour de 1150. Ces deux lemmes partagent un sémantisme extrêmement proche et une base étymologique commune (-RĒTRO), ce qui rend remarquable leur relatif retard conjoint, et ce d'autant qu'ils ne sont pas sans attestations non défectueuses avant 1137 et 1150, même si le nombre de textes reste peu élevé. La conjonction d'une forte concentration dans le temps de toutes les premières occurrences, meilleure que pour les autres catégories jusqu'ici étudiées, et de cette spécificité de « derriere » et « ariere » rend l'ensemble difficile à analyser. En effet, alors que le premier paramètre plaide pour une absence de diffusion du défaut dans le corps lexical, la diffusion précédant alors notre corpus, le second paramètre plaide pour une diffusion progressive, les items formés sur -RĒTRO n'étant atteints qu'en fin de queue. Nous ne sommes pas en mesure de trancher, dans l'absolu, entre ces deux lectures, et ces éléments sont synthétisés, avec leur incertitude, en (7.32).

(7.32) **Adv : diffusion lexicale.**

L'apparition extrêmement précoce des premières occurrences défectueuses semble plaider pour une absence de diffusion du défaut au sein du corps lexical. En revanche, le léger retard de « ariere » et « derriere » et leurs taux inférieurs semblent témoigner, mais de manière peu consolidée, de la fin du processus de diffusion lexicale.

Par ailleurs, lors du commentaire du graphique de relation entre fréquence d'emploi et taux de défaut (fig. 7.7), nous n'avons noté aucun effet

de la fréquence lexicale sur les taux de défaut de <e>, ni même d'effet de filtre provoqué par de plus faibles fréquences. Fréquence lexicale et force du défaut de <e> semblent ici tout à fait indépendantes, dans le sens où la fréquence ne peut prédire le taux de défaut. En revanche, nous avons noté en sec. 7.2.4 qu'une plus faible fréquence semblait exprimer la dimension stochastique (c'est-à-dire liée à l'introduction du facteur chance dans le calcul des probabilités) de l'effet du contexte, dans la mesure où ce sont les quatre lemmes dont la fréquence est la plus faible qui témoignent de taux de défaut déséquilibrés entre le contexte prévocalique et le contexte préconsonantique (cf. *infra*). Ces éléments sont repris en (7.33).

(7.33) **Adv : fréquence.**

Les données plaident pour une absence d'effet de la fréquence lexicale sur les taux non contextuels de défaut de <e>. En revanche, la fréquence semble influencer sur la dimension stochastique du défaut de <e> (cf. 7.36).

Enfin, nous avons noté en sec. 7.2.3.4 que, dans *Roland*, la succession du lemme « ensemble » par la préposition *od* était systématiquement associée à un défaut de <e> (*ensembl od*), tandis que « ensemble » était sans défaut de <e> dans tous les autres contextes, y compris prévocaliques. Si le nombre d'occurrences est somme toute restreint (17 occ. *ensemble od* et 6 formes *ensemble* suivies d'un autre terme), la distribution complémentaire est parfaite. La collocation *ensembl od* semble donc avoir atteint le plus haut degré de grammaticalisation, et fonctionner non plus comme un *type* mais comme un *token* c'est-à-dire comme une seule unité au niveau du traitement. Ce fonctionnement de *token* peut expliquer, par deux dimensions simultanées, la disparition du <e> du premier terme : d'une part, dans un *token*, la frontière entre les deux composants est abolie dans le traitement cognitif de l'unité (qui ne représente plus qu'un seul constituant indivisible : c'est la *fusion des constituants* de BYBEE 2006), et, d'autre part, les segments qui composent les *tokens* sont fragilisés par la vitesse de traitement accrue des *tokens*, ce qui résulte couramment en des suppressions de segments (REDUCING EFFECT : BYBEE 2000, 2002, 2006 ; BYBEE et SCHEIBMAN 1999). Ici, ces deux dimensions semblent tracer clairement une limite dans le continuum entre *types* et *tokens*, dans la mesure où l'abolition de la frontière de domaine et/ou la réduction des segments observables dans *ensembl od* ne s'applique pas aux autres occurrences prévocaliques. Cet effet est néanmoins strictement limité à *Roland*, et l'on ne peut donc savoir s'il représente une simple idiosyncrasie du scripteur, un trait anglo-normand archaïque, ou un trait plus général mais qui aurait échappé à la graphie dans le reste de ses occurrences. Ces éléments sont résumés en (7.34)

(7.34) **Adv : collocations.**

Dans *Roland*, la collocation grammaticalisée *ensembl od* est toujours en défaut de <e> et représente le seul contexte où le lemme « ensemble » est en défaut de <e>.

7.3.2 Le rôle du contexte

Nous avons vu en sec. 7.2.4 que si la statistique inférentielle repère un effet significatif du contexte préconsonantique sur la génération de défauts de <e>, cet effet n'est pas nécessairement pertinent sur le plan linguistique. En effet, l'effet est très faible ($t = -2,5; p < 0,05$),²⁴ particulièrement instable dans le temps et dans l'espace, et contre-intuitif sur le plan phonologique, impliquant plus de défauts de <e> entre deux consonnes que devant une voyelle. Nous proposons donc de considérer qu'il ne s'agit que d'un effet de bruit dans les données, sans intérêt pour notre analyse. Ces éléments sont repris en (7.35). Les valeurs reportées en (7.35) sont celles calculées sur les données en fig. 7.8, c'est-à-dire à l'exception des formes à 's adverbial' et du lemme « ensemble ».

24: Test de Student à variables dépendantes, tel qu'en fig. 7.8.

(7.35) **Adv : contexte.**

Hormis pour « ensemble » (cf. 7.34), l'effet statistique faible du contexte subséquent sur la génération de défauts de <e> ne semble pas pertinent sur le plan de l'analyse linguistique.

($t = -2,5; p < 0,05$; moy. __#C = 74%; moy. __#V = 68%)

Ce faible effet est particulièrement instable dans le temps et dans l'espace.

Nous avons pu observer lemme par lemme (table 7.6) que cette absence d'effet du contexte, valable pour la somme des huit lemmes, l'est également pour quatre lemmes individuels, mais que deux lemmes favorisent le défaut dans le contexte prévocalique, tandis que deux autres lemmes favorisent le défaut dans le contexte préconsonantique. Étant donné que ces quatre lemmes sont ceux dont la fréquence est la moins élevée (de 256 à 552 *vs* de 936 à 6027 pour les autres lemmes), qu'ils se divisent quant au contexte qui favorise le plus le défaut de <e> et que la situation des deux lemmes qui favorisent le défaut dans le contexte préconsonantique est difficile à analyser en termes phonologiques, nous avons conclu qu'il ne s'agissait là que d'un effet du hasard. La généralisation en (7.36) synthétise cette dimension stochastique du phénomène de défaut de <e> des adverbes.

(7.36) **Adv : lemmes et contexte.**

La moitié des lemmes dont la fréquence est la moins élevée (a) montre une sensibilité erratique au contexte (b). Cette différence n'est probablement due qu'à la dimension stochastique du phénomène en jeu, qui ne s'exprime qu'en-deçà d'un certain seuil de fréquence (cf. 7.33).

a) $F \leq 552$ *vs* $F \geq 936$ pour les autres lemmes,

b) « onque », « mar » : __#V favorise le défaut de <e>

« derriere », « desor » : __#C favorise le défaut de <e>.

Comme déjà évoqué pour les prépositions « sor » et « desor » (sec. 6.3.2), POPE (1966 [1934], p. 118, 218) et FOUCHÉ (1969 [1958], p. 513) avancent

POPE (1966 [1934]). *From Latin to Modern French with Especial Consideration of Anglo-Norman.*

FOUCHÉ (1969 [1958]). *Phonétique historique du français — Les voyelles.*

que le schwa final des adverbes « ariere », « derriere », « encore », « lors » et « ore » est *optionnel* (POPE 1966 [1934], p. 118) car ces formes étaient, originellement, pré-vocaliques, avant que la chute de leur <e> devant initiale vocalique n'entraîne la chute du schwa dans la forme lexicale. Pour les prépositions comme pour les adverbes, nos données ne permettent pas de confirmer cette hypothèse : non seulement le contexte préconsonantique est beaucoup plus courant que le contexte prévocalique en a.fr., mais le contexte préconsonantique est aussi associé à un taux de défaut plus élevé que ne l'est le contexte prévocalique. S'il est possible que l'antériorité de la chute de <e> devant initiale vocalique précède notre corpus, MEYER-LÜBKE (1890, p. 558-559) rejette cette hypothèse au moins pour l'adverbe « ore », dont il note que des textes anciens ne connaissent, devant initiale consonantique, que *or* (*Alexis, Gormont, Voyage de Charlemagne*).²⁵ Nous avons nous-même relevé (en sec. 7.2.3.2) que la première occurrence défectueuse de *lors* dans notre corpus se trouvait devant initiale consonantique dans *gorm*, c'est-à-dire *Gormont*, et avons également noté que dans *Roland* (sec. 7.2.3.1), le lemme « ore » est en défaut de <e> (*or*) dans 15 occurrences préconsonantiques, tandis qu'il n'apparaît avec son <e> (*ore*) que dans 5 occurrences préconsonantiques. Pour Meyer-Lübke, l'hypothèse d'une chute prévocalique qui aurait été rephonologisée comme une chute non contextuelle ne tient donc pas, et ce sont plutôt la fréquence élevée de ces adverbes et leur emploi grammatical qui expliqueraient la chute du <e>.

Si nous n'avons trouvé aucune corrélation entre les taux de défaut des lemmes et leur fréquence au sein de la catégorie des adverbes, il convient de noter que la fréquence d'emploi des adverbes est, avec celle de certaines prépositions, bien supérieure à celle que nous avons pu observer pour les noms et les adjectifs féminins singuliers. En effet, le nom sujet à défauts de <e> dont la fréquence est la plus élevée, « guerre », n'est attesté qu'en 529 occ., et l'adjectif féminin singulier sujet à défauts de <e> dont la fréquence est la plus élevée, « bon », n'est attesté qu'en 1405 occ., tandis qu'une préposition comme « sor » est attestée en 3590 occ.²⁶ et que l'adverbe « ore » est attesté par 6027 occ.²⁷ Cela n'est néanmoins pas vrai de tous les lemmes prépositionnels et adverbiaux, avec par exemple 263 attestations de « mar ». Néanmoins, au-delà de la fréquence attestée par les textes, il est fort probable que la fréquence d'emploi des items des catégories fonctionnelles (comme les prépositions) soit, dans une langue naturelle, bien supérieure à celle des catégories lexicales à inventaire plus large et ouvert. Si, sur ce point, la typologie des adverbes pose problème (EGF),²⁸ il nous suffit de noter que les adverbes entretiennent une porosité avec les prépositions encore plus vivace en a.fr. qu'en fr.mod. (PICOCHÉ et MARCHELLO-NIZIA 1994 [1989], p. 271; MARCHELLO-NIZIA 1999b, p. 120; MÉNARD 1988 [1973], p. 262), et que le classement d'un item en tant qu'adverbe découle du simple fait qu'il est invariable mais n'est ni une préposition, ni une conjonction, ni une interjection (RIEGEL, PELLAT et RENÉ 2009 [1994], p. 646). Par ailleurs, adverbes et prépositions partagent un certain nombre de fonctionnements phonologiques ; en fr.mod. par exemple, adverbes et prépositions sont soumis à la même restriction de monosyllabité pour déclencher la liaison consonantique obligatoire entre eux et le nom suivant (*très* [z] *ému* ; *en* [n] *Asie*), alors que les adjectifs ne sont pas soumis à cette restriction de monosyllabité pour le déclenchement de la liaison avec le nom suivant (*lointain* [n] *ami* ; NGUYEN et al. 2007, p. 9 ; ENCREVÉ 1983, p. 52.

MEYER-LÜBKE (1890). *Grammaire des langues romanes*.

25: Le texte, mais pas nécessairement l'édition, du *Gormont* cité par Meyer-Lübke correspond à **gorm** : ANON. (1125*). *Gormont et Isembart*. Dees : Angleterre ; compo. frc., ca 1125 ; ms. a.n., ca 1213 ; MS2 (BAYOT 1931) ; DEAF : **GormB**. *Voyage de Charlemagne à Constantinople*, compo. 2nd m. xii^e s. (source DEAF : **PelCharK**) ou 3^e q. xii^e s. (source : Arlima, https://www.arlima.net/uz/voyage_de_charlemagne.html).

26: Ce nombre ne concerne que les occurrences de « sor » annotées pos = "PRE".

27: Ce nombre ne concerne que les occurrences de « or » annotées pos = "ADV".

28: L'article « La paradigmatique : statut des catégories » de *L'Encyclopédie grammaticale du français* est collectif ; il est indexé sous BLANCHE-BENVENISTE et al. (2020 [2018]) dans la bibliographie.

RIEGEL et al. (2009 [1994]). *Grammaire méthodique du français*.

NGUYEN et al. (2007). « Detection of liaison consonants in speech processing in French: Experimental data and theoretical implications ».

ENCREVÉ (1983). « La liaison sans enchaînement ».

De ce point de vue, les adverbes, ou du moins certains d'entre eux, sont à rapprocher des mots fonctionnels et non des mots lexicaux. Ce rapprochement ne résout en rien la question, dans la mesure où certaines prépositions, au premier rang desquelles « de », montrent une sensibilité extrême au contexte (cf. sec. 6.2.4.1). Néanmoins, en accord avec les observations de Meyer-Lübke, et en l'absence de relevé d'occurrences chez Pope et Fouché, il convient de remettre en question le fondement empirique de l'assertion de ces deux auteurs.

7.3.3 Diachronie générale

Nous avons observé en sec. 7.2.2 que la dynamique diachronique générale, telle que représentée par la régression linéaire sur les graphiques de séries temporelles non contextuels en fig. 7.1, 7.2 et 7.3 *supra*, est à la hausse du taux de défaut de <e>. Cette dynamique se retrouve également dans chaque contexte isolé, comme l'illustrent les régressions linéaires en fig. 7.8.a-d *supra*. Néanmoins, cette observation résulte essentiellement d'un écrasement de la variation diatopique : l'investigation du rôle et du poids de l'anglo-normand dans cette dynamique révèle que cette hausse générale n'est vraie ni en anglo-normand ni sur le domaine continental, que ce soit devant initiale consonantique ou devant initiale vocalique. Cet effet contre-intuitif s'explique par la combinaison de deux caractéristiques des textes anglo-normands, illustrées en fig. 7.22 : ceux-ci sont majoritairement précoces et disposent de taux le plus souvent bien inférieurs à la moyenne.

Ainsi, le taux moyen de défaut de <e>, pondéré par le nombre d'occurrences des textes, est de 46% devant initiale consonantique en anglo-normand, alors qu'il est de 81% en a.fr. continental. Même chose devant initiale vocalique : le taux moyen anglo-normand est de 32%, alors qu'il est de 75% en a.fr. continental. Par ailleurs, la date moyenne des textes, c'est-à-dire la moyenne des dates de tous les textes, est antérieure de 33 à 35 ans en anglo-normand, étant donné la concentration des textes anglo-normands dans la seconde moitié du XII^e s.

Une inspection visuelle, pour ainsi dire naïve, des graphiques en fig. 7.22 permet de comprendre cette situation ; pour faciliter l'exposition, la figure a été annotée d'un encadré par graphique. Il s'agit de percevoir que les droites de régression linéaire à taux croissant vues auparavant résultent en réalité de l'équilibre fait par le modèle de régression entre le pôle anglo-normand à valeur faibles (de ~0,23% à ~85% devant initiale consonantique ; de 0% à ~40% devant initiale vocalique, cf. encadré (i) en fig. 7.22.a,c) dans la seconde moitié du XII^e s. et la masse plus ou moins constante de textes continentaux dont les valeurs dépassent ~70%. Mais, lorsque le pôle anglo-normand est soustrait aux données continentales, celles-ci changent d'équilibre, étant polarisées par une densité supérieure de textes à taux élevés à la fin du XII^e s. et au début du XIII^e s. (encadrés (ii) en fig. 7.22.b,d), et leur taux moyen est alors légèrement décroissant devant initiale consonantique, et presque constant devant initiale vocalique.

Aussi, il convient de ne pas considérer que les taux de défaut de <e> des adverbes sont croissants comme s'il s'agissait d'une tendance linguistique générale. Il s'agit en effet, comme cela pouvait être le cas dans les autres

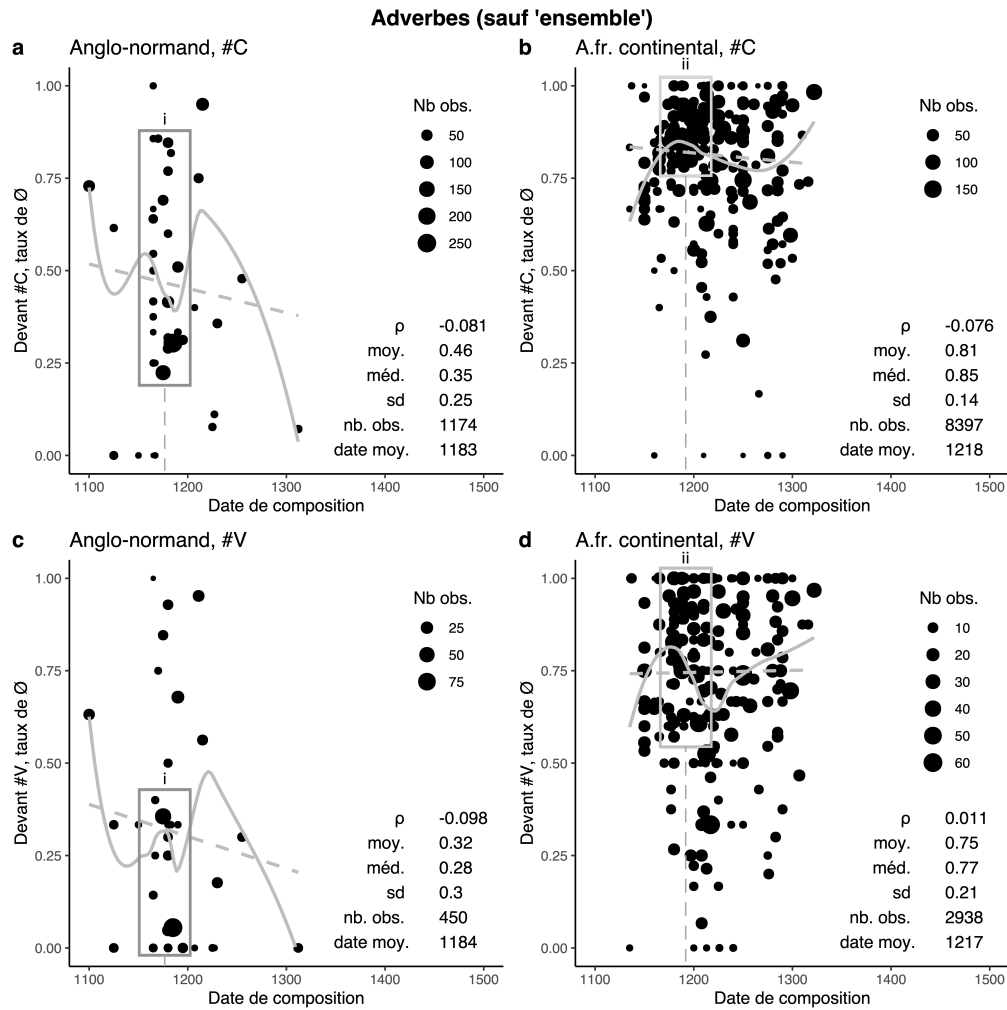


FIGURE 7.22 – Diachronie des taux de défaut de <e>, anglo-normand *vs* ancien français continental, adverbes. Les régressions locales en (a) et (c) peuvent être ignorées étant donné le faible nombre de textes hors de la seconde moitié du xii^e s.

parties du discours des inflexions temporaires des régressions locales, d'un effet diatopique lié à la spécificité de l'anglo-normand. Sur le continent, le taux de défaut de <e> est relativement stable, ce qui va dans le sens d'une alternance <e>/Ø qui ne se diffuse pas mais qui est présente, d'emblée, et se maintient telle quelle.

(7.37) **Adv : diachronie.**

Les taux de défaut de <e> sont constants. L'observation d'une tendance à la croissance est due à l'influence de l'anglo-normand (cf. 7.39).

7.3.4 Diatopie

Nous avons observé en sec. 7.2.5 une distribution diatopique marquée, en synchronie, par un arc de cercle partant de l'Angleterre, passant par le centre du domaine et aboutissant soit au Sud-Est du domaine (devant initiale consonantique) soit au Sud-Sud-Est du domaine (devant

initiale vocalique). Cet arc de cercle concentre les valeurs inférieures à la moyenne des taux régionaux, et il est plus large devant initiale vocalique que devant initiale consonantique. Comme en témoigne l'indice de prévalence de l'éllision, cet arc de cercle est aussi la partie du domaine où le contexte préconsonantique est substantiellement associé à plus de défauts de <e> que le contexte prévocalique. Hors de cet arc de cercle, les taux moyens régionaux s'étagent entre la moyenne et des valeurs de 94% de défauts devant initiale consonantique et 95% devant initiale vocalique, les valeurs les plus élevées étant classiquement rencontrées à la frontière Est et Nord-Est du domaine. À l'Ouest de l'arc de cercle des taux inférieurs à la moyenne, se trouvent de nombreuses régions non attestées par le corpus, aussi la délimitation occidentale de l'arc de cercle découle seulement de la Normandie, de la région Vendée, Deux-Sèvres et, devant initiale vocalique, de la Charente-Maritime.

Diachroniquement, cette situation synchronique ne commence à se définir clairement devant initiale vocalique que dans la première moitié du XIII^e s., tandis qu'elle semble se définir dès la seconde moitié du XII^e s. devant initiale consonantique. Si la situation se stabilise dans les deux contextes au XIII^e s., elle est donc encore mouvante au XII^e s. L'anglo-normand, s'il n'est attesté de manière satisfaisante que dans la seconde moitié du XII^e s., présente à toutes les époques des taux de défaut très bas, qui sont souvent le minimum de l'ensemble du domaine (cf. sec. suivante). Si l'Angleterre apparaît d'abord détachée du domaine continental, l'apparition de l'arc de cercle des taux de défaut les plus faibles au XIII^e s. et l'ouverture de cet arc de cercle sur la Manche par la région Somme, Pas-de-Calais place finalement l'Angleterre dans la continuité de celui-ci. Aussi, la continuité partielle entre l'Angleterre et la partie du domaine continental où les taux sont les plus faibles, continuité que l'on peut observer en synchronie, n'est pas originelle ; mais, dans la seconde moitié du XII^e s., cette absence de continuité peut être compensée par un certain mitage territorial qui autorise les apparitions éparses de taux bas sur le continent également.

(7.38) **Adv : diatopie.**

Un arc de cercle Angleterre - Centre - Sud-Est du domaine comporte tous les taux inférieurs à la moyenne. Cet arc est plus large devant initiale vocalique. Les taux les plus élevés se situent à l'Est du domaine.

Cette distribution diatopique n'est stable qu'à partir du XIII^e s.

7.3.5 Poids de l'anglo-normand et de son hétérogénéité

L'Angleterre apparaît isolée du domaine continental dès la première moitié du XII^e s., et le reste sur l'ensemble du corpus. Elle ne connaît pas, pour les adverbes, de variations diachroniques erratiques comme elle peut en connaître pour les adjectifs féminins singuliers ou pour les prépositions, comme illustré en table 7.7. Le domaine anglo-normand est donc, pour les adverbes, caractérisé par une stabilité plus forte que pour les parties du discours étudiées précédemment. Au contraire, une tendance cohérente à la baisse des taux de défaut se dégage, mais celle-ci est peu consolidée par le nombre de textes.

TABLE 7.7 – taux de défaut de <e> (Tx.Ø) devant initiale vocalique en anglo-normand, adjectifs féminins singulier, prépositions « de » et adverbes.

Période	Tx.Ø	Tx.Ø	Tx.Ø		Nb. textes
	AdjFÉMSg	PRÉP « de »	ADV		
	__#V	__#V	__#V	__#C	
(1) 1100-1150	8%	89%	32%	33%	4
(2) 1151-1200	21%	64%	33%	50%	34
(3) 1201-1250	5%	79%	28%	44%	6
(4) 1251-1322	37%	51%	15%	27%	2

L'anglo-normand est néanmoins caractérisé par une hétérogénéité synchronique marquée. Comme vu en fig. 7.22 *supra*, l'écart-type pondéré par le nombre d'occurrences de chaque texte est légèrement plus élevé en anglo-normand que sur le continent, que ce soit devant initiale consonantique (a.n. $\sigma = 0,25$; a.fr.cont. $\sigma = 0,14$) ou devant initiale vocalique (a.n. $\sigma = 0,3$; a.fr.cont. $\sigma = 0,21$). L'Angleterre diffère également du domaine continental par ses taux de défaut, avec cette fois-ci une différence plus nette : la moyenne des taux de défaut par texte, pondérée par le nombre d'occurrences de chaque texte, est de 46% devant initiale consonantique en a.n. (*vs* 81% en a.fr.cont.) et de 32% devant initiale vocalique (*vs* 75% en a.fr.cont.).

Cette différence entre l'Angleterre et le domaine continental est évidemment statistiquement significative ; elle est représentée pour le contexte prévoicalique en fig. 7.23 — sur cette figure, les mesures ne sont pas pondérées par le nombre d'occurrences des textes.

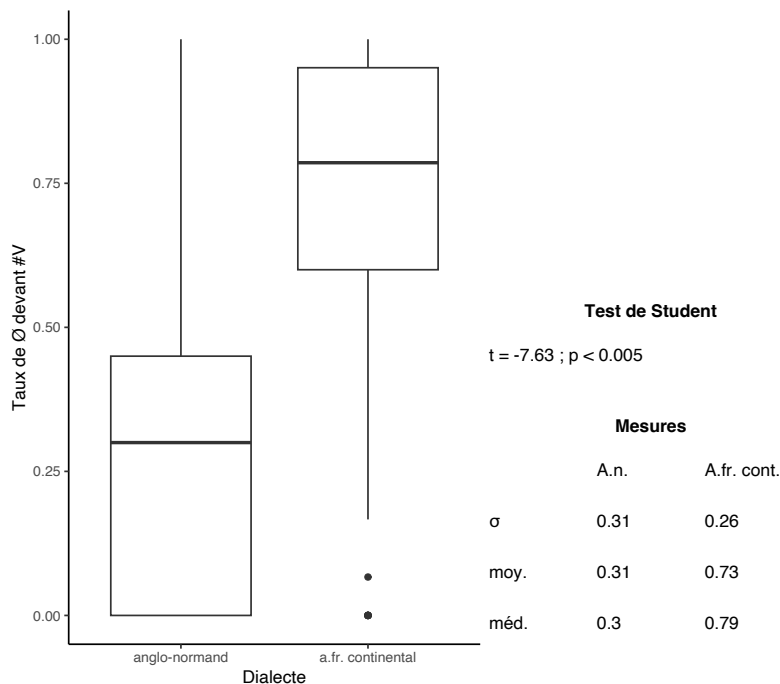


FIGURE 7.23 – Diachronie des taux de défaut de <e>, anglo-normand *vs* ancien français continental, adverbes (sauf « ensemble ») devant initiale vocalique.

Ainsi, devant initiale vocalique, le domaine anglo-normand est différent du domaine continental ($t = -7,63$) et cette différence est significative ($p < 0,005$). La spécificité de l'anglo-normand, ici, est plus forte que celle observée pour les adjectifs féminins singuliers ($t = 4,11$) et pour les prépositions ($t = -5,33$), et elle rejoint celle des prépositions dans les valeurs négatives (l'a.n. a moins de défauts que l'a.fr.cont.), par opposition aux adjectifs féminins singuliers, où la valeur- t était positive (l'a.n. avait plus de défauts que l'a.fr.cont.). Comme relevé en (7.37), c'est cette spécificité de l'anglo-normand qui est à l'origine du caractère croissant des régressions linéaires de l'ensemble du domaine, peu importe le contexte, alors qu'une fois Angleterre et continent séparés, les taux ne croissent plus avec le temps. Néanmoins, ici, l'a.n. ne saurait être responsable des oscillations temporaires du taux moyen des textes; comme représenté en fig. 7.22 *supra*, le français continental est lui-même marqué par un pic des taux de défaut de <e> avant 1200 devant initiale vocalique, et autour de 1200 devant initiale consonantique. Tout au plus, la présence de nombreux textes anglo-normands dans la seconde moitié du XII^e s. est responsable d'un amoindrissement des pics observés, ce qui est l'inverse de l'effet observé jusqu'à présent pour les autres catégories.

Enfin, il est néanmoins notable que, si cette différence implique une dispersion des données un peu plus étendue en anglo-normand, cette dispersion est sans commune mesure avec l'hétérogénéité observée pour les parties du discours étudiées précédemment. En effet, l'anglo-normand dispose d'un écart-type 3 fois plus élevé que celui de l'ensemble continental pour les adjectifs féminins singuliers (___#V : a.n. $\sigma = 0,23$; a.fr.cont. $\sigma = 0,08$)²⁹ et 3,5 fois plus élevé pour les prépositions (___#V : a.n. $\sigma = 0,32$; a.fr.cont. $\sigma = 0,09$).³⁰ Pour les adverbes, ce n'est donc pas tant par son hétérogénéité que par ses taux anormalement bas que se signale l'anglo-normand. L'ensemble de ces éléments est repris en (7.39) et (7.40). En (7.39), par souci de cohérence avec les généralisations présentées précédemment, nous n'indiquons que les mesures du contexte prévoicalique, et ce sont les mesures non pondérées telles que présentées en fig. 7.23.

29: Cf. fig. 5.6.

30: Cf. fig. 6.10.

(7.39) **Adv : spécificité de l'a.n.**

L'a.n. (a) se distingue des dialectes continentaux par (b) ses nombreux textes à faibles taux de défaut de <e>, mais (c) son hétérogénéité n'est qu'à peine supérieure à celle du domaine continental.

- a) ___#V : $t = -7,63$; $p < 0,005$;
- b) ___#V : moy. = 31% vs a.fr. continental moy. = 73%;
- c) ___#V : $\sigma = 0,31$ vs a.fr. continental $\sigma = 0,26$.

(7.40) **Adv : poids de l'a.n. dans les données.**

Étant donné (7.39), plus il y a de textes anglo-normands, moins les défauts de <e> (devant initiale vocalique comme consonantique) sont nombreux (cf. 7.37)

7.3.6 Place de la variation au sein des Adv

Si les défauts de <e> sont majoritaires au sein de nos neuf ou huit lemmes (68%, ou 75% à l'exception du lemme « ensemble »), ils sont minoritaires

au regard de l'ensemble des occurrences adverbiales, ne touchant que 5,8% des formes (sec. 7.2). Néanmoins, comme expliqué en sections 7.2.1 et 7.3.1, ces huit ou neuf lemmes représentent une part non négligeable des adverbes qui disposent d'un <e> sur certaines occurrences, dont le <e> est de droit, et qui disposent d'une fréquence absolue satisfaisante. Par ailleurs, ce taux de variantes sans <e> sur l'ensemble des lemmes de la catégorie est bien supérieur à celui observé pour les noms (<0,01%; sec. 4.2) et adjectifs féminins singuliers (0,57%; sec. 5.2), et au contraire très proche du taux observé pour les prépositions (5,99%).

En revanche, le taux de défaut de <e> des adverbes sur les seules occurrences des lemmes retenus (68% ou 75%) est bien supérieur à ce même taux pour les prépositions, qui n'est que de 16%. Cela est notamment dû au fait que les défauts de certaines prépositions sont conditionnés par le contexte prévocalique (seuls « sor » et « desor » ont des défauts devant initiale consonantique), lequel est bien moins courant que le contexte préconsonantique en a.fr. Il est sur ce point notable que les occurrences prépositionnelles, ou du moins annotées comme telles, des lemmes « sor » et « desor » se comportent donc comme des adverbes de plein droit et non comme des prépositions. Ces éléments sont repris en (7.41), dont les valeurs sont celles qui font exception du lemme « ensemble », étant donné la circonscription de ses défauts à *Roland*.

(7.41) **Adv : poids de la variation.**

Rapportés à l'ensemble des Adv, les défauts de <e> sont globalement minoritaires (a), mais sont majoritaires pour les lemmes concernés (b) et ceux-ci représentent une part conséquente des lemmes susceptibles de connaître un tel défaut (cf. 7.31).

a) Défauts : 6% des occ. d'Adv,

b) Défauts : 75% des occ. des 8 lemmes retenus.

7.4 Conclusion

Dans ce chapitre, nous avons étudié la distribution des défauts de <e> des adverbes dans le corpus du NCA. Dans l'introduction, après une note sur la pérennité de ces défauts, au sein d'un environnement plus large d'alternances <e>/Ø pour les adverbes (sec. 7.1.1), nous avons présenté les exclusions de données (sec. 7.1.2) et les formes relemmatisées (sec. 7.1.3). Nous avons ensuite présenté les résultats de l'analyse que Comparalem nous a permis de mener (sec. 7.2), et avons proposé une brève typologie préliminaire des lemmes et formes sujets à défauts de (sec. 7.2.1). Après avoir brièvement abordé les principales caractéristiques de la distribution de ces défauts de <e> en diachronie (sec. 7.2.2), nous avons développé la spécificité de chaque lemme par ordre d'apparition dans le corpus, et avons traité de la collocation *ensembl od* (sec. 7.2.3), avant de traiter de la distribution des défauts de <e> selon le contexte phonologique (sec. 7.2.4), en diatopie (sec. 7.2.5), et à l'interface entre diachronie et diatopie (sec. 7.2.6). Enfin, nous avons procédé à une montée en généralisation, qui tient lieu de *discussion* des résultats observés (sec. 7.3).

Nous avons ainsi pu mettre en évidence (sec. 7.3.1) qu'une part non négligeable des lemmes qui réunissaient les conditions pour un défaut de <e> étaient bien atteinte par ce défaut (7.3.1), que c'était déjà le cas dès le début de notre corpus (7.32), que les taux de défaut n'étaient pas corrélés à la fréquence des items (7.33) et que seul « ensemble » montrait une association stricte entre défaut de <e> et construction figée (*ensembl od*), et ce uniquement dans *Roland* (7.34). En sec. 7.3.2, nous avons établi que le contexte phonologique n'avait pas d'effet linguistique fort sur les défauts de <e> de nos lemmes pris dans leur ensemble (7.35). En revanche, les lemmes dont la fréquence est la plus faible montrent une préférence pour les défauts dans l'un ou l'autre contexte, mais la possibilité que ce soit aussi bien le contexte préconsonantique que le contexte prévocalique qui soit associé au plus de défauts de <e> semble indiquer qu'il s'agit là d'un effet stochastique et non d'un comportement proprement phonologique (7.36).

En ce qui concerne la diachronie (sec. 7.3.3), nous avons montré que la tendance générale à la hausse des taux de défaut n'était qu'un artéfact dû à l'intersection de la diatopie et de la diachronie, et en particulier au positionnement de l'anglo-normand au sein de la diachronie du corpus. De ce fait, il s'agit plutôt de considérer que les taux de défaut de <e> sont relativement stables dans le temps si l'on neutralise ce biais diatopique (7.37).

De façon plus générale, la distribution diatopique des données met en avant la formation, progressive et aboutie uniquement au ^{xiii}e s., d'un arc de cercle partant de l'Angleterre et rejoignant le Sud ou le Sud-Est du domaine en passant par le centre de celui-ci ; cet arc de cercle concentre les taux inférieurs à la moyenne (7.38).

Nous avons ensuite complété l'étude de la spécificité de l'anglo-normand en notant que, pour les adverbes, ce dialecte ne connaissait plus les alternances fort marquées de taux de zéro selon les périodes qu'il connaît pour d'autres parties du discours, mais en notant que l'Angleterre demeurerait fortement différenciée du domaine continental, que ce soit par ses taux bas ou par son hétérogénéité interne (7.39). Ces facteurs expliquent que la présence de textes anglo-normands abaisse les taux de défaut de <e>, devant initiale vocalique comme devant initiale consonantique (7.40).

Enfin, nous avons pris du recul pour observer le poids de la variation de nos huit lemmes adverbiaux au sein de la catégorie des adverbes (sec. 7.3.6), et avons observé que les taux de défaut des adverbes sont bien supérieurs à ceux des noms et adjectifs féminins singuliers (sur l'ensemble des catégories comme sur les seuls lemmes retenus), mais que si le taux de défaut sur l'ensemble de la catégorie est équivalent à celui des prépositions, le taux de défaut sur les seuls lemmes retenus (c'est-à-dire sur les lemmes sujets à un véritable défaut de <e>) est bien plus élevé pour les adverbes que pour les prépositions (7.41).

8.1 Introduction

Ce chapitre est dédié à l'analyse de la variation <e>/Ø et <V>/Ø des articles de l'ancien français; étant donnée la fréquence très forte des articles, nous avons choisi de nous focaliser sur cette sous-catégorie plutôt que sur la catégorie des déterminants en général. Tout comme en français moderne, les articles connaissent l'élision (*le + empereor* → *l'empereor* : « l'empereur »),¹ y compris avec d'autres voyelles que schwa (*li + emperere* → *l'emperere*; *la + amie* → *l'amie*),² et cette élision peut revêtir un caractère systématique.

Après avoir introduit les principaux éléments pertinents concernant la formation et le fonctionnement des articles en a.fr. (sec. 8.1.1) et expliqué la chaîne de traitement des données adoptée (8.1.2), nous présentons les résultats pour l'ensemble des articles sujets à l'élision (sec. 8.2). Cette présentation comporte une typologie des articles vis-à-vis de leurs formes où une voyelle finale fait défaut (sec. 8.2.1), puis détaille les données concernant les articles dont l'élision est systématique (sec. 8.2.2), ceux dont elle est optionnelle (sec. 8.2.3) et enfin ceux pour lesquels elle est prohibée (sec. 8.2.4). Enfin, en sec. 8.3, nous regroupons l'ensemble des informations collectées pour obtenir une vue d'ensemble des défauts vocaliques parmi les articles.

8.1.1 Les articles de l'ancien français

L'article représente, vis-à-vis du latin, une innovation morphosyntaxique : alors que la détermination est rare en latin, elle devient courante en a.fr., avant de devenir dominante en fr.mod. L'ancien français n'est donc pas encore dans la situation du français moderne, où la détermination du nom est la règle. Reprenant CARLIER et GOYENS (1998) et GOYENS (1994), CARLIER (2020, GGHF, p. 664-665) compare la version latine du *De Inuentione* de Cicéron, sa traduction en a.fr. par Jean d'Antioche à la fin du XIII^e s. et sa traduction en fr.mod. : dans la version latine, 13% des noms sont pourvus d'un élément externe assurant la détermination, tandis que dans la version médiévale, ce sont 59% des noms qui sont déterminés, et que dans la version moderne, 84% des noms sont déterminés. Sur les 60% de noms déterminés en a.fr., 44% sont déterminés par un article défini, 3% par l'indéfini *un(s)* et 12% par un autre déterminant. L'article indéfini *un(s)* est encore rare en a.fr. ; sa fréquence triple de l'a.fr. (3% chez Jean d'Antioche) au fr.mod. (11%). Enfin, le partitif, quasiment absent de l'a.fr., ne se développe qu'à la fin du XIV^e s. et au début du XV^e s. (GGHF *id.* p. 665).

Les articles de l'ancien français se résument ainsi à deux items : l'article défini et l'article indéfini. Le premier provient du démonstratif latin *ILLE*, -A et le second du numéral *UNUS*, -A ; leur paradigme est donné en (8.1). En latin comme en a.fr., ces deux termes se déclinent en cas et

1: Pex. *Roland*, v. 954, 3678; cf. sec. 8.2.2.

2: *L'emperere* : cf. sec. 8.2.3; *l'amie* : p.ex. e*lid* p. 20, *cou* p. 8.

elid : MARIE DE FRANCE (ca 1165). *Lais*. Compo. NO, ca 1165; ms. a.n., ca 1275; MS (MARIE DE FRANCE 1947).

cou : ANON. (1210*). *Courtois d'Arras*. Dees : Hainaut; compo. art., ca 1210; ms. pic., ca 1285; cr1 (FARAL 1922).

CARLIER et al. (1998). « De l'ancien français au français moderne : régression du degré zéro de la détermination et restructuration du système des articles ».

GOYENS (1994). *Émergence et évolution du syntagme nominal en français*.

CARLIER (2020). « Les articles définis, indéfinis et partitifs ».

en genre ; néanmoins, la déclinaison de l'a.fr. *li* au MASC CSsg n'est pas étymologique, et la déclinaison du féminin fait défaut dès le t.a.fr., les formes du cas sujet s'étant alignées sur celles du cas régime. En (8.1), les flèches barrées '→' désignent ces cas où la descendance est indirecte, la finale ayant été modifiée par une analogie.

(8.1) Articles de l'a.fr. (formes canoniques)

	MASC	FÉM.
(a) Article défini		
.i. CS sg	ILLE → <i>li</i>	ILLA → <i>la</i>
.ii. CR sg	ILLUM → <i>le</i>	ILLAM → <i>la</i>
.iii. CS pl	ILLĪ → <i>li</i>	ILLAE → <i>les</i>
.iv. CR pl	ILLOS → <i>les</i>	ILLAS → <i>les</i>
(b) Article indéfini		
.i. CS sg	ŪNUS → <i>uns</i>	ŪNA → <i>une</i>
.ii. CR sg	ŪNUM → <i>un</i>	ŪNAM → <i>une</i>
.iii. CS pl	ŪNI → <i>un</i>	ŪNAE → <i>unes</i>
.iv. CR pl	ŪNOS → <i>uns</i>	ŪNAS → <i>unes</i>

À l'exception de l'article indéfini masculin, toutes ces formes disposent d'une voyelle finale qui est de droit et qui provient d'une finale atone latine. Dans la série des articles définis, l'aphérèse a privé le mot de sa première voyelle, ce qui assure automatiquement la sauvegarde de sa finale atone, devenue seule voyelle du mot. Pour les indéfinis féminins, c'est la présence d'un -A latin qui assure la retenue d'une voyelle finale atone, que l'aphérèse ait eu lieu (articles définis) ou non (articles indéfinis). Les formes en (8.1) ne sont que les plus courantes ; il existe une variation diatopique du timbre des voyelles pour tous les cas sauf le CRsg, dont les principales variantes sont indiquées en (8.2), version adaptée et complétée du tableau de CARLIER (2020, GGHF, p. 667).

(8.2) Variantes des articles définis

	MASC	FÉM.
(a) CS sg	<i>le</i> (a.n.)	<i>le</i> (a.n.) <i>le, li</i> (pic., wall., norm.)
(b) CR sg	<i>lo</i> (norm. arch.) <i>lou</i> (pic.) <i>lu, lo</i> (a.n.)	<i>le</i> (a.n., pic., wall.)
(c) CR pl	<i>lé</i> (a.n., pic.)	<i>lé</i> (a.n., pic.)

Cette variation tient notamment, en picard, à l'emploi d'une forme épicène pour le CRsg (*le*), qui s'est étendue au FÉMCS (le féminin n'a plus de cas) voire a entraîné une ré-introduction de la déclinaison (*li* au FÉMCSsg) par alignement général du féminin sg sur le masculin sg (GOSSEN 1970, p. 121). Si la forme *li* pour le FÉMCSsg ne semble pas attestée en a.n., l'usage de formes épicènes est un trait non seulement du picard mais de toute une partie des dialectes du Nord-Est, concernant également la moitié Nord du normand (comprise dans l'étiquette 'normanno-picard'), le wallon (GOSSEN 1970, p. 121) et l'anglo-normand (SHORT 2007, p. 45). Ces formes *le, li* au féminin ne sont pas exclusives des formes *la* (qualifiées de « franciennes » par Gossen), dont la quantité va croissante et finit par dominer à partir du xiv^e s. Parallèlement, des formes à voyelle autre que <e> sont conservées au MASC CRsg en normand (*lo*), en picard (*lou*)

GOSSEN (1970). *Grammaire de l'ancien picard*.

SHORT (2007). *Manual of Anglo-Norman*.

et en anglo-normand (*lu, lo*); ces formes, tenues pour archaïques et absentes des autres dialectes, se maintiennent au moins jusqu'au xiv^e s. en anglo-normand, mais parfois dans des constructions plus ou moins figées (SHORT 2007, p. 126).

Aussi, le timbre des articles définis montre une certaine variation, et ne correspond pas nécessairement à un schwa. Néanmoins, comme sera démontré *infra*, la voyelle de l'article n'a pas besoin d'être un schwa pour être élidée. La situation est ici analogue à celle du pronom régime féminin de 3sg *la* et des formes atones ou non spécifiées de 1sg *je, jo, jou* (etc.), qui s'élient sans que leur voyelle ne soit un schwa (cf. 9.21). Néanmoins, et encore une fois comme pour les pronoms personnels, le fait que des voyelles autres que schwa s'élient ne veut pas dire que toutes les voyelles subissent l'éliision. Dans son point sur l'éliision des articles définis, CARLIER (2020, GGHF, p. 668) indique en effet que si les CR *le* et *la* s'élient, le MAscCSsg *li* ne s'élide que « occasionnellement » et le MAscCSpl *li* « ne s'élide pratiquement jamais ». Nos données confirment ces observations, tout en fournissant une quantification précise des phénomènes en question.

8.1.2 Traitement des données

Comme pour les pronoms, notre chaîne de traitement des données, depuis TXM jusqu'à Comparalem, est ici différente de ce qui a été proposé pour les parties du discours précédentes. En effet, un rapide examen des occurrences a révélé que les seules formes d'articles définis courantes (F > 50) et dépourvues de consonne finale sont *la, li, le, lo, lou, lu, ly* et *l*. De même, pour les articles indéfinis, les formes courantes (F > 50) sont *un(e), vn(e), una* et *ung*. Pour accélérer le traitement des données et isoler d'emblée des erreurs d'annotation, nous avons donc effectué les requêtes CQP en ciblant les occurrences de ces formes annotées comme des déterminants; nous nous passons donc des lemmes ici.³

Comme pour les pronoms, les requêtes sont ici rédigées de manière à opposer les formes défectueuses (*l, un, ung*) et les formes non défectueuses, que leur voyelle finale soit représentée par <e> (*le, une*) ou par un autre graphème (*la, li, lo, una*, etc.). Enfin, pour des raisons inhérentes à cette chaîne de traitement particulière et au nombre très élevé d'occurrences, les occurrences n'ont pas été systématiquement nettoyées manuellement avant ou pendant leur traitement par Comparalem; aussi certains chiffres et certaines sorties graphiques sont discutées et rectifiées dans le texte.

3: En pratique, les lemmes des articles définis ont été redéfinis sur la valeur fake-lemma tandis que les articles indéfinis n'avaient déjà qu'un seul lemme et ne nécessitaient pas d'opération de relemmatisation pour être traités par Comparalem.

8.2 Présentation des résultats

Nous travaillons ici sur les articles définis qui sont susceptibles de connaître un défaut de leur voyelle finale, ce qui ne concerne donc qu'une partie des articles. Pour les articles définis, sont concernés le masculin au CSsg (*li*), CRsg (*le*) et CSpl (*li*) ainsi que le féminin, aux CS et CR sg (*la*). Pour les articles indéfinis, cela concerne uniquement les féminins singuliers (*une*). La table 8.1 présente les taux de défaut de voyelle finale atone pour les articles retenus, selon le cas et le genre. Elle précise, par ordre de colonne, le nombre total d'occurrences (dans les contextes retenus),

puis le nombre d'occurrences (avec et sans voyelle finale atone) devant initiale consonantique et le taux de formes défectueuses parmi ces formes, avant d'indiquer les mêmes données pour les occurrences prévocaliques. Sur cette table, les articles définis féminins singuliers figurent deux fois ; la première ligne les concernant comprend toutes les formes annotées comme féminines (y compris les occurrences dialectales du type *le*, *li*) tandis que la seconde ne comprend que les formes canoniques *la* et *l*.

TABLE 8.1 – Quantité de formes et taux de formes en défaut de <e> pour les articles.

	Genre	Cas	Nb.	_#C		_#V	
				Nb.	Tx_Ø	Nb.	Tx_Ø
DÉF	MASC	CS sg	29 346	24 940	<0,01%	4406	33%
		CR sg	31 345	22 894	<0,01%	8451	99%
		CS pl	8379	6747	<0,01%	1632	1%
	FÉM ^a	CS-R sg	51 436	42 741	<0,01%	8695	97%
	FÉM ^b	CS-R sg	48 327	39 862	<0,01%	8465	99%
INDÉF	FÉM	CS-R sg	7568	6317	1% ^c	1251	9%

a. Toutes occurrences annotées comme féminines, y compris *li*, *le*.

b. Sans formes épiciques et/ou analogiques (*la* et *l* uniquement).

c. Révisé à 0,3% en sec. 8.2.3.2.

Ces données confirment donc la remarque de CARLIER (2020, GGHF, p. 668) mentionné *supra*, avec une élision presque systématique de l'article défini au féminin singulier sujet comme régime (*la*) et au masculin singulier régime (*le*), une élision non systématique au masculin singulier sujet (*li*) et une quasi-absence d'élision au masculin pluriel sujet (*li*). L'autrice notait également (p. 672-673) que le <e> de l'article indéfini féminin *une* peut chuter devant voyelle, « sans que le phénomène soit systématique » ; nos données montrent que cette absence de systématisme correspond à un taux d'élision assez bas (9%). Enfin, pour tous ces articles, Comparalem a détecté une quantité plus ou moins faible (de <0,01% à 1%) de formes défectueuses *l* et *un* devant initiale consonantique ; nous verrons *infra* que ces formes, en plus d'être fortement marginales pour quatre catégories sur cinq (< 0,01%), résultent d'erreurs d'annotations et de cas d'enclise analogues à ceux étudiés pour les pronoms personnels (cf. sec. 9.2.2.1).

Étant donnés ces résultats, dans ce qui suit, nous ne reprenons pas chaque catégorie dans une section dédiée, mais détaillons les résultats en regroupant les cas par leurs taux de défaut devant initiale vocalique, selon la typologie esquissée dans la section suivante.

8.2.1 Typologie

Au vu des taux représentés en table 8.1, il convient de retenir que certains articles sont sujets à une élision presque systématique (*le* CRsg et *la* CS-Rsg), que d'autres sont sujets à une élision facultative (*li* CSsg et *une* CS-Rsg), tandis que l'article défini masculin pluriel régime *li* ne s'élide pas. Comme indiqué en (8.3), cette typologie doit traiter différemment les articles définis féminins analogiques *li* picards et wallons. En effet, si l'inclusion des formes *le* pour *la* ne change pas les taux d'élision, la forme *li* pour le CSsg féminin ne s'élide pas systématiquement, se comportant comme au masculin ; c'est cette forme qui est responsable des 2% d'écart

entre le taux d'élision de l'article féminin singulier *la* (99%) et celui des articles féminins singuliers *la*, *le* et *li* (97%). Il n'est pas possible, au demeurant, de calculer le taux d'élision de la forme *li* du CSsg féminin, sa forme élidée étant indissociable de celles des formes non analogiques *la* et *le* (l).

(8.3) Typologie en fonction des taux de défaut devant #V

(a) Élision systématique

- .i. DÉFMASC sg CR $le \rightarrow l / _ \#V$ (99%)
- .ii. DÉFFÉM sg CS-R $la (le) \rightarrow l / _ \#V$ (99%)

(b) Élision facultative

- .i. DÉFMASC sg CS $li \rightarrow l / _ \#V$ (33%)
- .ii. INDÉFFÉM sg CS-R $une \rightarrow un / _ \#V$ (9%)
- .iii. DÉFFÉM sg CS $li \rightarrow l _ \#V$ (?)

(c) Élision prohibée

- .i. DÉFMASC pl CS $li \rightarrow li / _ \#V$ (> 99%)

Comme pour les pronoms personnels, cette typologie fondée sur les taux d'élision ne recoupe que partiellement une typologie qui serait fondée sur l'identité phonologique de la voyelle défectueuse. En effet, non seulement *la* (FÉM_{sg}) et *le* (MASC_{CRsg}, FÉM_{sg} pic.)⁴ s'élident systématiquement et n'ont pas la même voyelle, mais *une* (FÉM_{sg}) et *li* (MASC_{CSsg}, FÉM_{CSsg} pic.) ne s'élident pas systématiquement alors que *une* se termine en <e> pour schwa comme *le*, et, surtout, que *li* singulier peut s'élider de manière optionnelle alors que *li* pluriel (MASC_{CS}) ne s'élide pas. Il est donc, ici aussi, impossible de prédire sur un fondement phonologique quel article est sujet à l'élision et quel article ne l'est pas. Enfin, alors que pour les pronoms personnels nous avons proposé une typologie par différence entre les taux d'enclise et les taux d'élision, les taux de défaut devant initiale consonantique sont ici extrêmement faibles (< 0,01%) sauf pour l'indéfini *une*, pour lequel nous n'avons pourtant pas trouvé dans la littérature de mention d'enclise. Dans ce qui suit, les résultats sont donc présentés en fonction de la typologie fondée sur le degré d'élision avancée en (8.3).

4: Par simplicité, nous ne précisons que l'origine picarde, sans rappeler à chaque fois que la forme concerne également l'anglo-normand, une partie du domaine normand et le wallon.

8.2.2 Articles à élision systématique : *le*, *la*

Comme indiqué en table 8.1 *supra*, l'élision semble systématique pour l'article défini singulier au CS régime masculin (*le*) et aux cas sujet et régime féminins (*la*), alors que les défauts devant initiale consonantique sont marginaux. Les graphiques de séries temporelles en fig. 8.1, 8.2 et 8.3 en pages suivantes représentent les taux de défaut vocalique de ces articles, avec deux figures pour le féminin, selon que les formes dialectales *le* et *li* sont prises en compte (fig. 8.2) ou non (fig. 8.3). Si ces données comportent des erreurs d'annotation, comme le montrent la plupart des exemples d'enclise en sec. 8.2.2.2, le caractère systématique de l'alternance fait que la présence marginale de défauts d'annotation dans les données n'est pas problématique; peu importe que l'on doive retrancher une quantité plus ou moins conséquente d'occurrences, la situation d'ensemble reste la même.

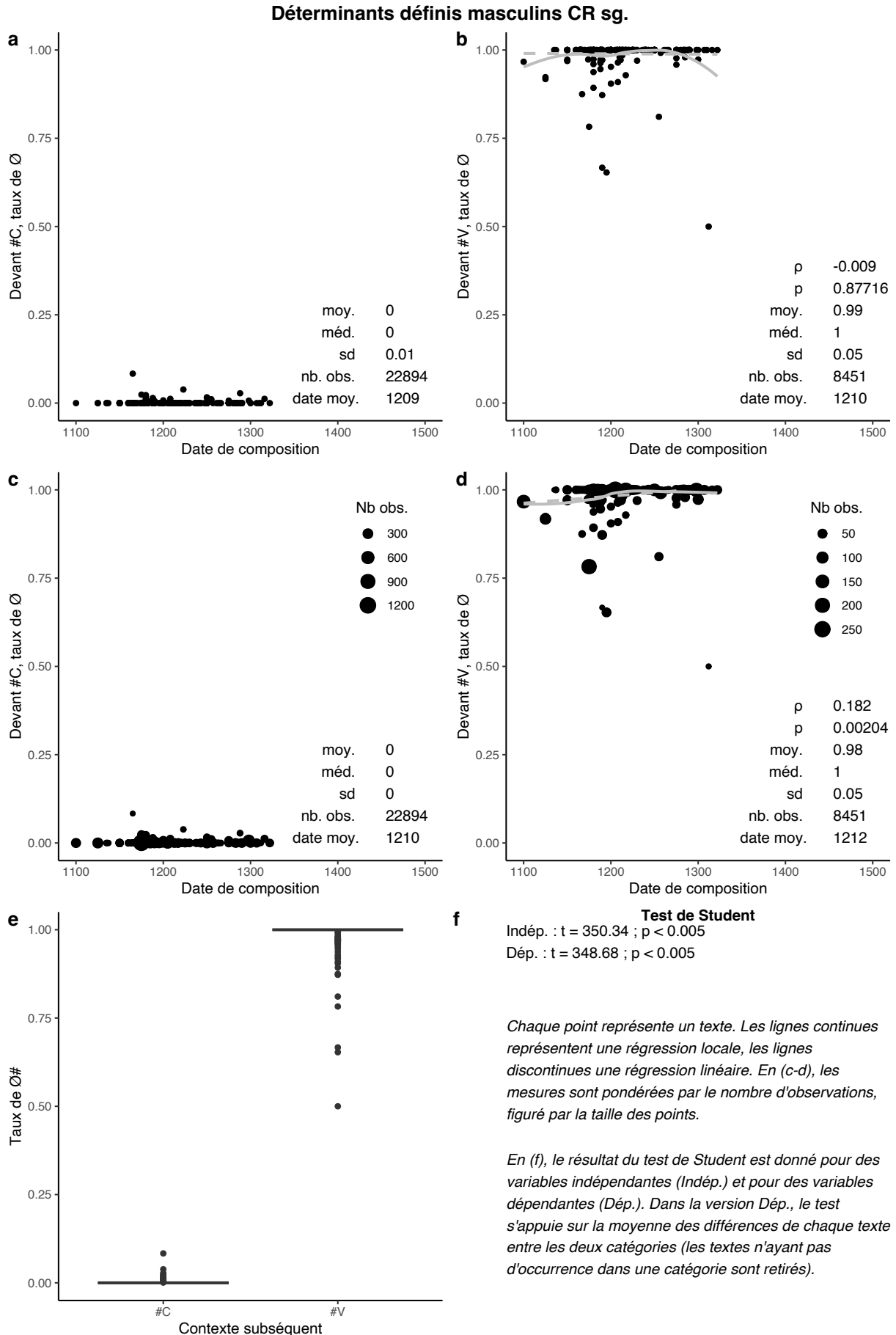


FIGURE 8.1 – Distribution contextuelle des variantes de articles définis masculins singuliers au cas régime en défaut de <e>.

8.2.2.1 Défauts devant #V

Devant initiale vocalique, l'élision est systématique : elle concerne 99% des occurrences de l'article défini singulier CR masculin *le* (fig. 8.1.b,d), et 97% des occurrences de l'article défini singulier féminin (CS et CR; fig. 8.2.b,d). Concernant l'article féminin, nous avons noté *supra* qu'il disposait de formes dialectales *le* (CS et CR), et que, sous l'influence de ces formes dialectales épïcènes, avait pu se développer également un CS *li* formé sur le masculin. Dans l'ensemble du corpus, 3906 formes *le* et *li* sont annotées comme étant des déterminants féminins, avec plus de formes *le* (2829 occ.) que de formes *li* (1077 occ.) — ce qui est attendu étant donné que *le* peut fonctionner au féminin pour les deux cas alors que *li* féminin analogique n'est attendu qu'au CS. Dans la mesure où l'article masculin CS *li* ne connaît pas une élision systématique (sec. 8.2.3), nous avons appliqué deux traitements au jeu de données de l'article défini féminin singulier, l'un comprenant toutes les formes annotées comme telles (fig. 8.2), et l'autre ne comprenant que les formes *la* et *l*, à l'exclusion des formes *le* et *li* (fig. 8.3).

Il ressort de ces manipulations qu'une partie non négligeable des formes non défectueuses devant initiale vocalique est due aux formes *li* : le retrait de ces occurrences fait remonter le taux de défaut devant #V de 97% à 99%. En revanche, le retrait des occurrences *le* devant initiale vocalique ne change pas considérablement la moyenne, qui demeure à 97%. Aussi, il convient de considérer que les articles définis féminins singuliers ont le même taux d'élision que l'article défini masculin CR *le*, et que c'est uniquement la présence de formes *li* pour le FÉMCS, par analogie au MAscCS, qui fait baisser le taux d'élision des articles définis féminins. Nous retenons donc le taux de 99% affiché en fig. 8.3, au détriment du taux de 97% affiché en fig. 8.2, en considérant que *li* analogique au FÉMCS_{sg} est exceptionnel par sa nature analogique et est traité comme le CS masculin et non comme le CS féminin ou le CR masculin.

Les cas d'absence de défaut ne représentent qu'environ 1% des occurrences devant initiale vocalique, avec 119 occurrences *le* au masculin (sur 8451 occ.) et 96 occurrences *la* ou *le* au féminin (sur 8695 occ.). Quelques cas peuvent être expliqués par un critère spécifique : en (8.4), le mot qui suit l'article, *osberc*, peut disposer d'un <h> initial et, étant d'origine germanique (FEW : frq. **halsberg*), celui-ci correspond à un 'h aspiré' ; le cas du mot *osberc* était déjà discuté en sec. 6.2.3, à propos de l'occurrence prépositionnelle *e elme ne d osberc* de *Roland*. La même chose vaut pour *le elme* (FEW frq. **helm*) en (8.5). Ces occurrences où un 'h aspiré' serait responsable de la non-élision du déterminant ne représentent néanmoins que 6 occurrences pour le masculin (5 avec *osberc*, 1 avec *elme*) et aucune pour le féminin.

- (8.4) *tot son glaive li empeint parmi l escu e par le osberc*
« son glaive l'atteint tout entier à travers son écu et son haubert »
(chro, p. 21, trad. pers.)
- (8.5) *li quens le fiert tant vertuosement*
tresqu al nase l tut le elme li fent⁵
« le comte le frappe avec une telle violence
qu'il lui fend tout le heaume jusqu'au nasal »
(rolandox, p. 101, v. 1644-1645, trad. SHORT (1990))

chro : BENOIT DE SAINTE-MAURE (1174*).
Chronique des ducs de Normandie. Ms. [Tours, Bibl. Munic. 903]; Dees : [Vendée, Deux-Sèvres]; compo. poit., ca 1174; ms. a.n., ca 1217; CR2 (BENOIT DE SAINTE-MAURE 1951-1967); DEAF : BenDucF.

5: SHORT (1990) corrige *helme*.

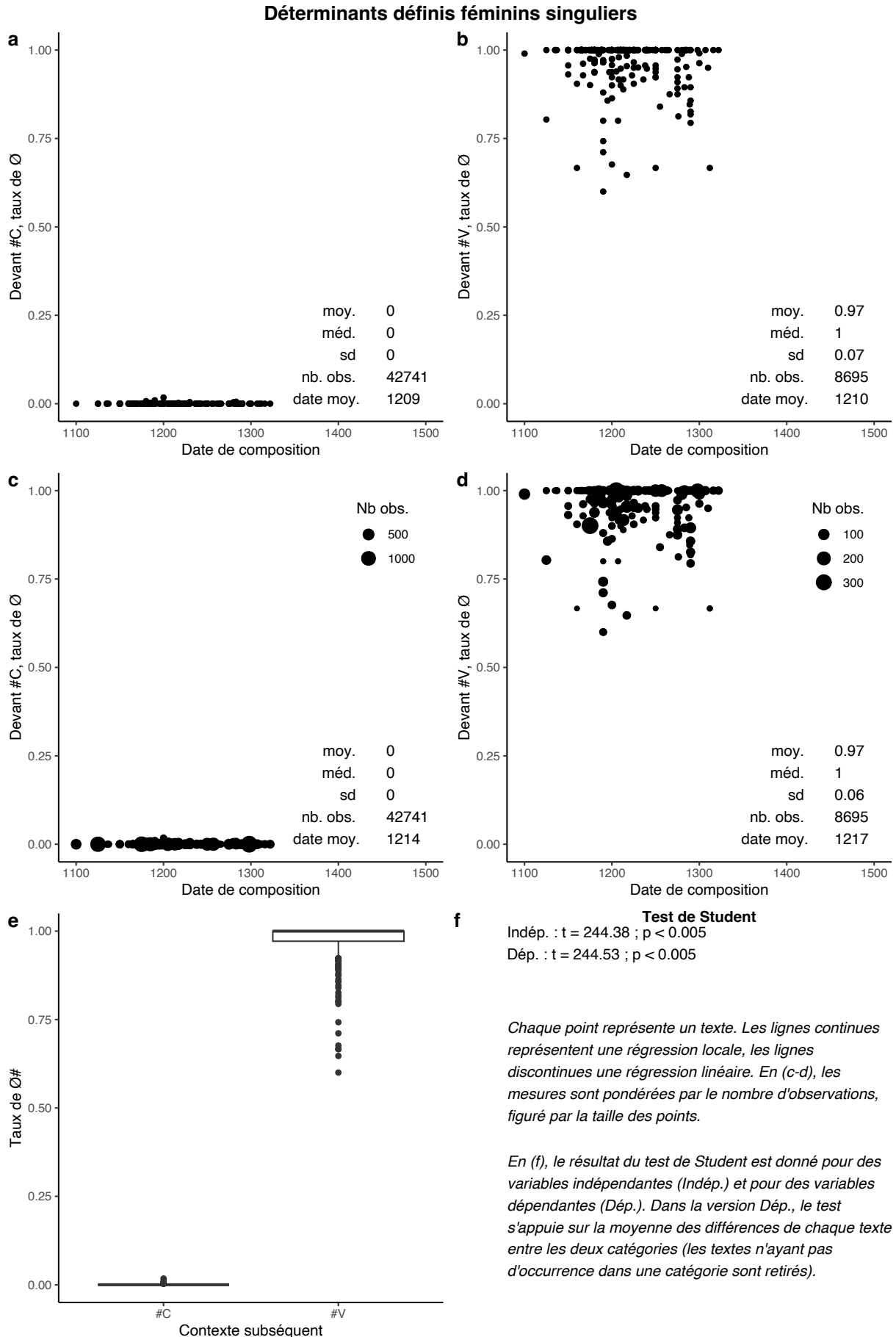


FIGURE 8.2 – Distribution contextuelle des variantes de articles définis féminins singuliers en défaut de <V>.

Déterminants définis féminins singuliers (formes 'la' et 'l' uniquement)

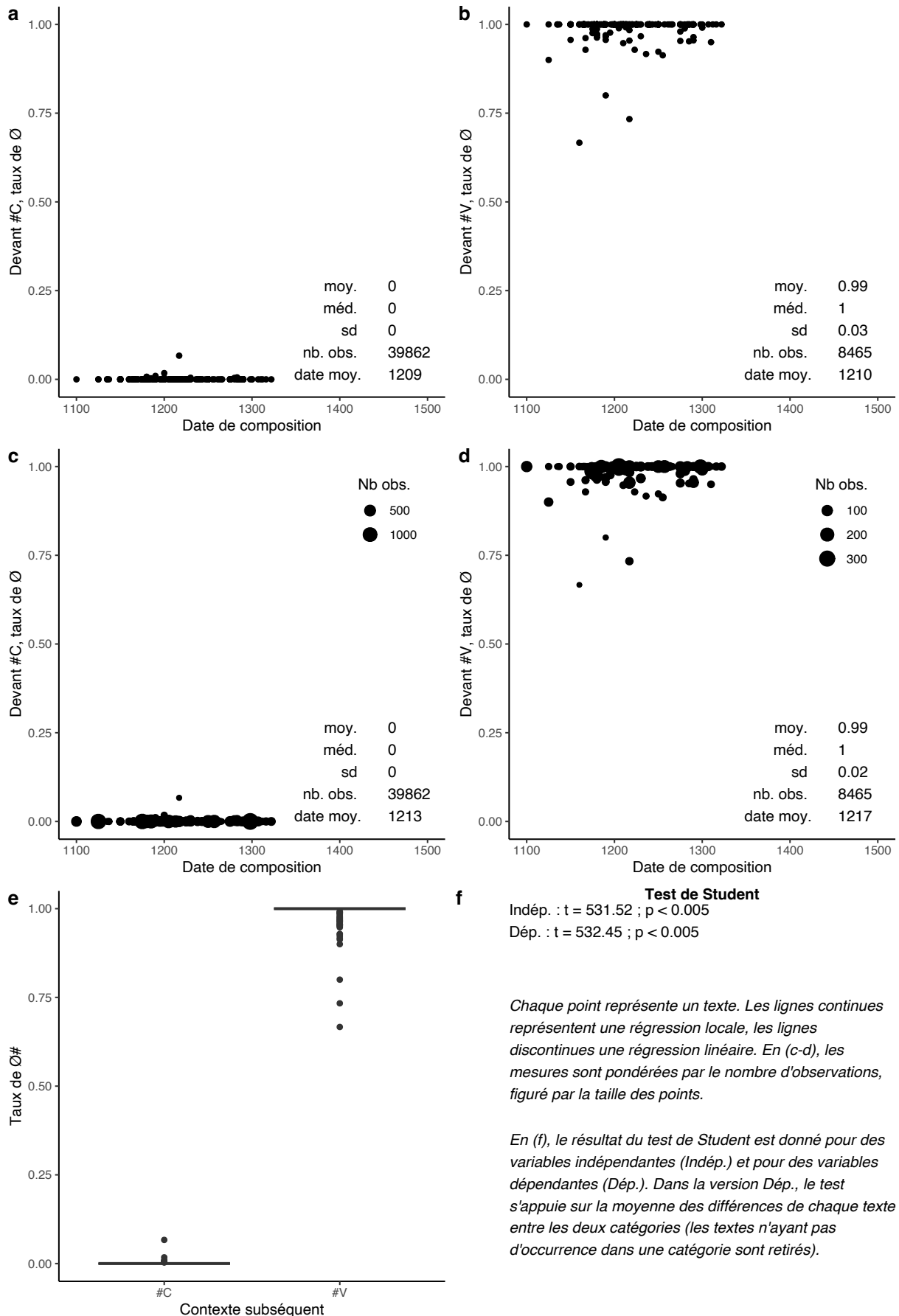


FIGURE 8.3 – Distribution contextuelle des variantes de articles définis féminins singuliers (formes *la* et *l* uniquement) en défaut de <a>.

Par ailleurs, si NYROP (1899, p. 228-231) remarque que l'élision est surtout évitée devant des noms propres, ce n'est le cas d'aucune de nos occurrences ici. Si quelques occurrences peuvent être suspectées d'un défaut d'annotation et si 10 occurrences tiennent à un chiffre transcrit *yy*, l'écrasante majorité de ces contre-exemples correspond bien à une structure *déterminant + nom* sans particularité apparente. On peut seulement relever que de nombreux exemples proviennent de textes liés à la religion (8.6 - 8.8), sans que ce soit une règle pour autant.

- (8.6) *li sires guarded les estranges le orfenin e la vedve*
« le Seigneur garde les étrangers, l'orphelin et la veuve »
(psautier, p. 120, trad. pers.)
- (8.7) *es paroles de la ewangel deués prendre le confort et la doctrine de uostre saluement*
« Dans les paroles de l'évangile vous devez trouver le réconfort et la doctrine de votre rédemption »
(carem, p. 16, trad. pers.)
- (8.8) *deuant tute sa gent gehi le auenture*
« devant toute sa troupe il [le roi] admit l'aventure »
(edmund, p. 25, trad. pers.)

Il faut donc admettre que le caractère systématique de l'élision de l'article défini singulier au cas régime masculin et aux cas sujet et régime féminins admet 1% de contre-exemples, du moins au niveau graphique.

8.2.2.2 Défauts devant #C

Pour ces articles, les cas de défauts vocaliques devant initiale consonantique sont extrêmement rares (< 0,01%). Sur les 39 862 occurrences de l'article masculin régime *le* devant initiale consonantique, 35 sont défectueuses (*l*). Sur ces 35 occurrences, 28 cas représentent une enclise, après *a* (8.9), *ne* (8.10), *ki* (8.12), *que* (8.13), *si* (8.14), *de* (8.15) et *contre* (8.16);⁶ sur ces 28 cas, l'annotation semble défectueuse dans la plupart des cas (il s'agit d'un pronom et non d'un article, comme tous les exemples *infra* sauf 8.9 et 8.11); dans quelques cas seulement il s'agit bien de l'article défini masculin singulier au cas régime (8.9, 8.11).

- (8.9) *quant gregoire a l duc ueu*
« quand Grégoire a vu le duc »
(pap, p. 29, trad. pers.)
- (8.10) *ie sui grosse de uif enfant ne l plus or mais celer auant*⁷
« je porte un enfant, je ne peux plus le cacher »
(pap, p. 5, trad. pers.)
- (8.11) *li vallez oi la nouele et l conseil*
« le valet a entendu la nouvelle et le conseil »
(deusamS, p. 3, trad. id.)
- (8.12) *ki l numer sace*⁸
« qui sache le nommer »
(edmond, p. 8, trad. id.)

NYROP (1899). *Grammaire historique de la langue française*.

psautier : ANON. (1125*). *Livre des psaumes*. Dees : nil; compo. a.n., ca 1125; ms. a.n., ca 1150; CR2 (MICHEL 1876), DEAF : PsCambrM.

carem : ANON. (1217*). *Sermons de carême*. Dees : Wallonie; compo. liég., ca 1217; ms. liég., ca 1225; MS2 (PASQUET 1888); DEAF : SermCarP.

edmund : DENIS PIRAMUS (1195*). *La Vie Saint Edmund le Roi*. Dees : [Angleterre]; compo. a.n., ca 1195; ms. a.n., ca 1325; MS2 (DENIS PIRAMUS 1935); DEAF : EdmK.

6: Tous ces cas sont enregistrés en tant qu'enclises dans les grammaires et dictionnaires; l'enclise, rare, avec *contre* est enregistrée par POPE (1966 [1934], p. 323).

pap : ANON. (1175*). *La vie du pape Saint Gregoire*. Ms. A1; Dees : nil; compo. Nord, ca 1175; ms. tour., ca 1238; MS1 (SOL 1977) DEAF : SGreg_a¹S.

7: *L* est annoté *pos* = "DET_def_obj_-masc_sg" alors qu'il s'agit ici vraisemblablement d'un pronom. C'est le cas des 7 occurrences *ne l*.

deusamS : MARIE DE FRANCE (1165*). *Les Deus Amanz*. Ms. S; Dees : nil; compo. a.n., ca 1165; ms. frc., ca 1300; MS (transcr. inédite sans transcripteur indiqué).

8: *Idem*; à moins que l'infinitif ne soit substantivé (*qui maîtrise la nomination*), il s'agit d'un pronom. L'autre occurrence *ki l* est clairement pronominale.

edmond : MATTHIEU PARIS (1255*). *La vie de saint Edmond*. Dees : nil; compo. a.n., ca 1255; ms. a.n., ca 1280; MS2 (J.-A. BAKER 1929); DEAF : SEdmCantB.

- (8.13) *mais de ce n est mais tenue la coustume des anciens, ja soit ce que l du pueple des gos et des alains et des unains*⁹
« la coutûme des anciens n'est plus respectée, que ce soit celle du peuple des Goths, des Alains ou des Unains (?) »
(artch, p. 21, trad. *id.*)
- (8.14) *si l pren si l jet*¹⁰
« et il le prend et le jette »
(reis, p. 262, trad. *id.*)
- (8.15) *puis prist de l norir grant conrei*¹¹
« puis il lui fit donner un grand repas »
(pap, p. 35, trad. *id.*)
- (8.16) *gent contra gent o ton cors contre l son*¹²
« groupe contre groupe, ou ton corps contre le sien »
(alia, p. 15, trad. *id.*)

Pour les articles féminins singuliers (CS comme CR), sur les 42 741 occurrences préconsonantiques, seules 10 sont en défaut vocalique. Certaines correspondent également à des cas d'enclise (5 occ.), et une erreur d'annotation semble également en cause, ces occurrences étant pronominales et, pour la plupart, masculines et non féminines (8.17 - 8.18).

- (8.17) *contient le livre de l que [...]*
« son livre [le livre de lui] expose que... » ou « le livre (la bible) dit à ce sujet que... »
(abe, p. 51, trad. pers.)
- (8.18) *ains ne l garda arme ne dras*
« jamais il ne lui resta ni arme ni vêtement »
(fablesM, p. 30, trad. *id.*)

Ces occurrences enclitiques, pour la plupart mal annotées, sont également accompagnées d'occurrences sans support d'enclise précédant le *l*; ces occurrences, pour problématiques qu'elles soient, sont si rares qu'on ne sait s'il faut leur faire un sort linguistique ou les considérer comme des *lapsus calami* ou éditoriaux, d'autant que nous avons déjà pu remarquer avec les pronoms des erreurs de transcription dans le corpus lui-même (sec. 9.2.2.1). Dans l'ensemble, si l'on cumule le fait que les cas d'enclise sont rarissimes, concernent pour la plupart un pronom et non un article, et qu'il peut exister à la marge des erreurs d'édition, il semble qu'il faille bien considérer que le défaut vocalique des articles *le* (MascCRsg) et *la* (FémCS-Rsg) sont exceptionnels devant initiale consonantique.

Le contexte joue donc, pour ces articles, un rôle fondamental dans les défauts vocaliques. Ce rôle, tel que mesuré par le test de Student à variables indépendantes en fig. 8.1.f, 8.2.f et 8.3 est le plus fort mesuré jusqu'à présent (MascCR : $t = 350$; Fém dont *le*, *li* : $t = 244$; Fém *la* uniquement : $t = 532$; $p < 0,005$ dans tous les cas). L'élimination des occurrences féminines *le* et surtout *li* devant initiale vocalique provoque donc une amélioration sensible de l'effet du contexte; ce qui ne s'exprimait que par 2% de différence en termes de taux moyens (de 97% à 99%) s'exprime par un doublement de la valeur- t (de 244 à 532). Cela vient confirmer plus fortement le rôle des articles féminins sujets *li* analogiques dans l'absence d'élision devant initiale vocalique. Étant donné le caractère presque systématique des taux observés pour ces articles, il n'est pas

9: *Idem.*

artch : JEAN DE MEUN (1288*). *L'art de chevalerie*. Dees : Marne; compo. frc., ca 1288; ms. pic., 1340; CR2 (JEAN DE MEUN 1897); DEAF : JMeunVégR.

10: Le premier *l* est annoté pos = "DET_ - def_obj_masc_sg" alors qu'il s'agit du pronom et non du déterminant; le second *l* est d'ailleurs bien annoté pos = "PRO_ - pers_obj_3_masc_sg".

reis : ANON. (1175*). *Les quatre livres des Rois*. Ms. M; Dees : nil; compo. a.n., ca 1175; ms. a.n., ca 1190; MS (prob. CURTIUS 1911).

11: *Idem.*

12: *Idem.*

alia : ANON. (1185*). *Roman d'Alexandre*. Ms. A; Dees : Vendée, Deux-Sèvres; compo. pic. et occ. orient., ca 1185; ms. poit., ca 1225; MS (ANON. 1937); DEAF : AlexArsL.

abe : JEAN DE MEUN (1288*). *Traduction de la première épître de P. Abélard*. Dees : région parisienne; compo. frc., ca 1280; ms. Paris, ca 1395; MS3 (JEAN DE MEUN 1934).

fablesM : MARIE DE FRANCE (1180*). *Fables*. Ms. M; Dees : nil; compo. NO, ca 1180; ms. wall., ca 1300; MS (transcription inédite).

nécessaire d'étudier leur distribution diatopique, qui est relativement homogène une fois les formes *li* du féminin éliminées.

8.2.2.3 En somme

Les articles masculins singuliers CR *le* et féminins singuliers CS et CR *la* et *le* subissent une élision presque systématique, avec 99% d'occurrences *l* devant initiale vocalique. Le 1% de formes non élidées devant initiale vocalique contient certes quelques cas particuliers et quelques erreurs d'annotation, mais il témoigne surtout de la possibilité de ne pas élider graphiquement l'article, de manière marginale, et sans qu'aucun critère ne permette de caractériser dans l'absolu ces occurrences. Devant initiale consonantique, les cas de défauts sont extrêmement rares (< 0,01%) et concernent des erreurs d'annotation et quelques cas d'enclise.

8.2.3 Articles à élision facultative : *li* (CSsg), *une*

Trois articles sont concernés par une élision en apparence optionnelle : l'article défini masculin singulier au cas sujet *li*, l'article indéfini féminin singulier *une* (cas sujet et régime) et l'article défini féminin dialectal *li* pour le cas sujet. Comme indiqué en table 8.1 *supra*, le premier connaît 33% de défauts vocaliques (*l*) devant initiale vocalique et le second en connaît 9%. Quant au troisième, nous ne pouvons pas mesurer ses défauts, dans la mesure où sa forme défectueuse *l* est impossible à distinguer de la forme défectueuse du féminin canonique *la* et de l'épicène dialectal *le*; la résistance partielle à l'élision de *li* FÉMCSsg est uniquement indiquée par le fait que les formes *li* sont responsables des deux tiers des cas de résistance à l'élision de l'article féminin, comme indiqué dans la section précédente. Pour *li* MASCCSsg et *une*, les figures en 8.4 et 8.6 (page suivante et p. 315) reproduisent la présentation habituelle des résultats, avec les graphiques de séries temporelles (a-d), les boîtes à moustaches (e) et le test de Student (f).

La comparaison entre ces deux figures montre que la situation de *li* MASCCS et celle de *une* FÉMCS-R sont fort différentes : l'effet du contexte est beaucoup plus marqué pour *li* ($t = 17,1$) que pour *une* ($t = 5,7$; $p < 0,005$ dans les deux cas), tout en demeurant sans commune mesure avec l'effet noté pour *le* et *la* (MASC CRsg et FÉMCS-Rsg). En effet, alors que *li* ne connaît pratiquement pas de défauts devant initiale consonantique, *une* en connaît un certain nombre (1%), ce qui est renforcé par le fait que *li* connaît bien plus de défauts devant initiale vocalique (33%) que *une* (9%). Ces chiffres, particulièrement pour *une*, devront néanmoins être révisés après inspection des occurrences.

8.2.3.1 Défauts devant #V

L MASC CSsg. Comme indiqué en fig. 8.4.b,d, l'article défini masculin CS est réduit à la forme *l* dans en moyenne 33% de ses occurrences prévoicales par texte. La courbe de régression locale non pondérée en fig. 8.4.b est moins stable que la courbe pondérée par le nombre d'occurrences des textes en fig. 8.4.d, ce qui indique un effet global de la taille des textes : ce sont les textes qui contiennent peu d'articles définis

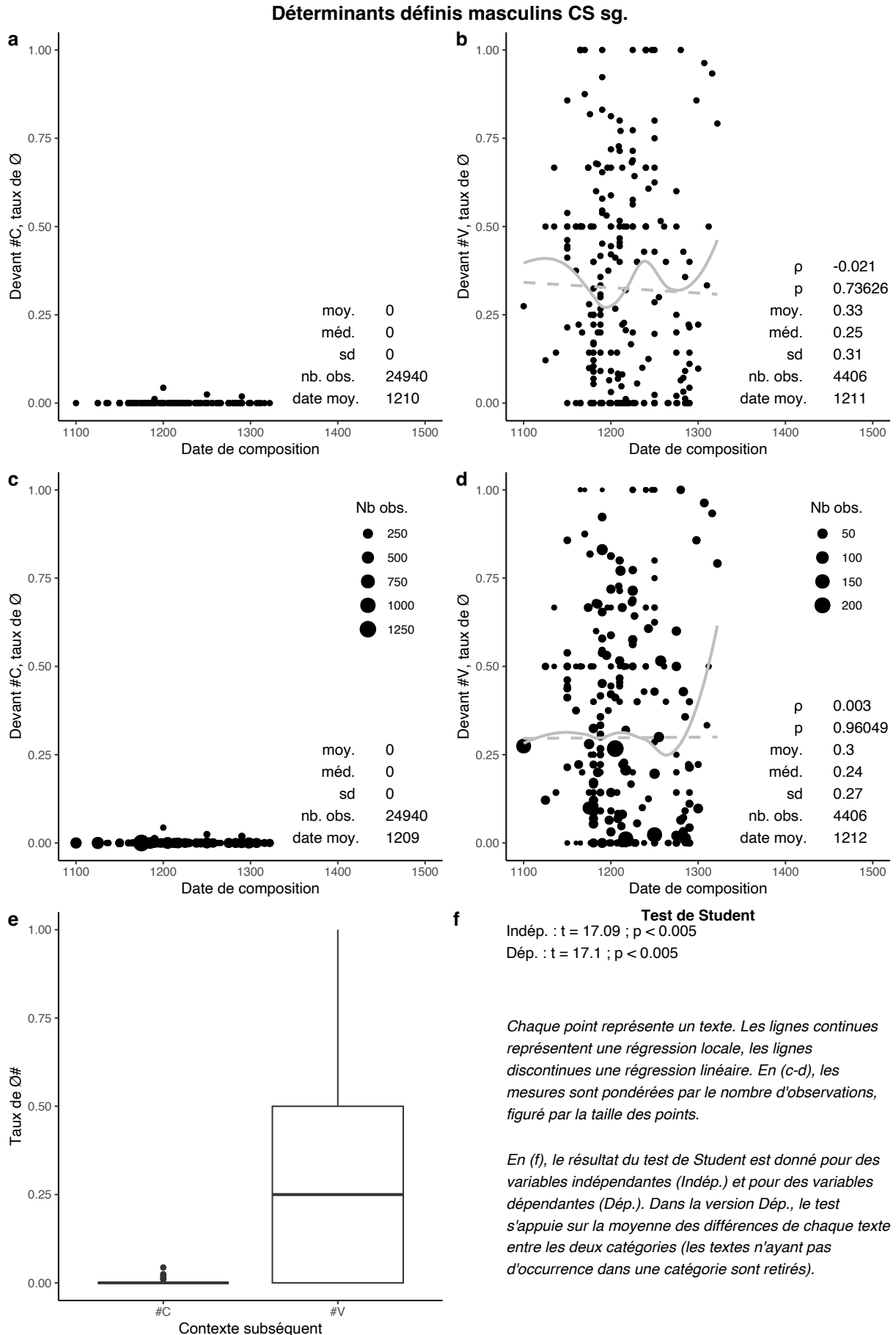


FIGURE 8.4 – Distribution contextuelle des variantes de articles définis masculins singuliers au cas sujet en défaut de <i>.

masculins singuliers CS qui ont les taux les plus irréguliers. Que les taux soient pondérés on non, la médiane est inférieure à la moyenne (méd. = 25%; méd. pond. = 24%), et indique que plus de la moitié des textes ne pratiquent le défaut vocalique sur l'article défini masculin CSsg que sur une occurrence sur quatre ou moins. Comme indiqué par la boîte à moustaches en fig. 8.4.e, un quart de nos textes ne pratique même aucune élision de *li* CSsg devant initiale vocalique. Néanmoins, ces textes ne correspondent à aucune unité diatopique ou diachronique donnée, et semblent se répartir de manière homogène dans le temps et l'espace.

Ce type de défaut est présent dès *Roland*, comme illustré en (8.19). Dans *Roland*, ces défauts vont de pair avec d'autres occurrences où l'élision n'est pas transcrite. Dans les cas où l'élision n'est pas transcrite, celle-ci peut être nécessaire à l'isosyllabisme du vers (comme pour *jo*, cf. sec. 9.1.1.1), mais parfois ce n'est pas le cas. Dans notre article utilisant le PAM pour découvrir les vers irréguliers de *Roland* (PREMAT et POGGIO [à paraître]), nous avons détecté 10 occurrences du syntagme *li emperere(s)* au CSsg où l'hémistiche contenant ce syntagme est hypermétrique, ce qui appelle l'élision métrique de *li* en *l*. Trois exemples en sont donnés en (8.20) - (8.22). À chaque fois, dans son édition plutôt interventionniste sur la syllabation à des fins métriques, SHORT (2007) soit édite *li* en *l* soit retranche un autre mot de l'hémistiche pour rétablir le bon nombre de syllabes métriques, tandis que l'édition du NCA (SEGRE 1971) et celle de la BFM (MOIGNET 1972a) respectent le ms., tel que reproduit en fig. 8.5, et éditent bien *li emperere*. Pour ces trois exemples, la lecture métrique à adopter est donc celle qui correspondrait à la graphie *l emperere(s)*, et non à la transcription *li emperere(s)*.

- (8.19) *vers dulce france chevalchet l emperere*
« Vers France la douce chevauche l'empereur. »
(rolandox, v. 706, trad. SHORT 1990)
- (8.20) *seignurs barons, dist li emperere carles*¹³
« « Seigneurs barons », dit l'empereur Charles »
(rolandox, v. 180, trad. *id.*)
- (8.21) *francs chevalers dist li emperere carles*¹⁴
« « Chevaliers francs », dit l'empereur Charles »
(rolandox, v. 274, trad. *id.*)
- (8.22) *e li empereres en est [e] ber e riches*¹⁵
« L'empereur en est célébré et puissant »
(rolandox, v. 2354, trad. *id.*)

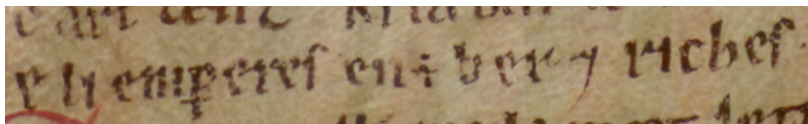


FIGURE 8.5 – *Roland*, ms. d'Oxford, fol. 42v, pénultième ligne (v. 2354). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

Short applique également l'élision à des contextes où elles ne nous semble pas nécessaire à la métrique; ainsi, en (8.23), il corrige *li emperere* en *l'empereur*, probablement parce qu'il n'admet pas de coupe épique après la sixième position métrique. En admettant *a priori* (c'est-à-dire à défaut de preuve du contraire) que les coupes épiques existent également

PREMAT et al. ([à paraître]). « Restituer des fragments de phonologie avec le PAM (*Programme d'Analyse Métrique*) : l'élision de schwa et des voyelles pleines dans les monosyllabes fonctionnels en très ancien français ».

13: SHORT (1990) corrige en *l'empereur*.

14: *Idem.*

15: L'ajout de [e] est de notre fait. SHORT (1990) conserve *li empereres* mais supprime le *e* à l'initiale du vers; il rajoute aussi un *e* devant *ber*. Le NCA (SHORT 1990) et la BFM (MOIGNET 1972a) lisent bien *e li empereres*; la BFM rajoute également un *e* devant *ber*.

après la sixième position métrique dans *Roland* (PREMAT et POGGIO [à paraître]), le second hémistiche formé par *li emperere* fait bien 4 syllabes métriques¹⁶ et correspond au patron métrique 4/6, dont il respecte le nombre et la position des syllabes métriques. Aussi nous ne cherchons pas à corriger ce vers — le caractère hypermétrique de la septième syllabe est indiqué par les chevrons précédant la coupe, notée '/'.

- (8.23) *mult fierement cheval<chet> / li emperere*¹⁷
 « L'empereur chevauche terrible et fier »
 (rolandox, v. 3316, trad. *id.*)

Au total, dans *Roland*, ce syntagme ARTDÉFCSSg + *emperere* requiert, au niveau métrique, une élision dans 27 occurrences. Cette élision est bien transcrite sur le ms. dans 17 occurrences (*l emperere(s)*), mais 10 occurrences du type de (8.20) - (8.22) ne notent pas l'élision (*li emperere(s)*) et seule la métrique permet de la restituer. Au contraire, c'est le maintien de *li* qui est nécessaire à la métrique dans 47 occurrences (8.24), dont seules 2 occurrences sont élidées (*l emperere* au lieu de *li emperere* : 8.25).

- (8.24) *esveillez est li emperere carles*
 « Charles l'empereur est réveillé »
 (rolandox, v. 2846, trad. SHORT 1990)
- (8.25) *co dist marsilie carles l emperere*¹⁸
 « Marsile leur dit : « L'empereur Charles »
 (rolandox, v. 2755, trad. SHORT 1990)

Ainsi, les élisions de *li* CSsg, relativement stables dans le temps si l'on se fie au jeu de données pondérées en fig. 8.4.d, sont bien attestées dès *Roland*, et la métrique propose des attestations non graphiques. Si la métrique s'oppose rarement à une élision transcrite dans la graphie (2 occ. sur 47), elle propose en revanche d'appliquer 10 élisions qui ne sont pas transcrites, ce qui viendrait sous-entendre que l'élision graphique se situe en-deçà de la productivité de l'élision métrique, et que le récitant de *Roland* saurait, pour les besoins de l'isosyllabisme, appliquer des élisions qui sont possibles mais ne sont pas notées.

Un FÉMSG. Les cas de défaut de <e> de l'article indéfini féminin singulier *une* devant initiale vocalique représentent 102 occurrences. Il faut d'emblée noter qu'un peu moins de la moitié de ces cas sont dus à une erreur d'annotation en genre, soit que le mot puisse avoir les deux genres (8.26), soit qu'il ne puisse être que masculin (8.27). Mais, dans un peu plus de la moitié des occurrences relevées, le mot déterminé est bien féminin et la forme attendue serait *une* et non *un*, comme c'est le cas en (8.28) et (8.29) par exemple.

16: Le -e de *emperere* est extramétrique, étant la finale atone d'un paroxyton dans une assonance dite féminine.

17: SHORT (1990) corrige en *l'empereur*.

18: SHORT (1990) corrige en *li emperere*.

Déterminants indéfinis féminins singuliers

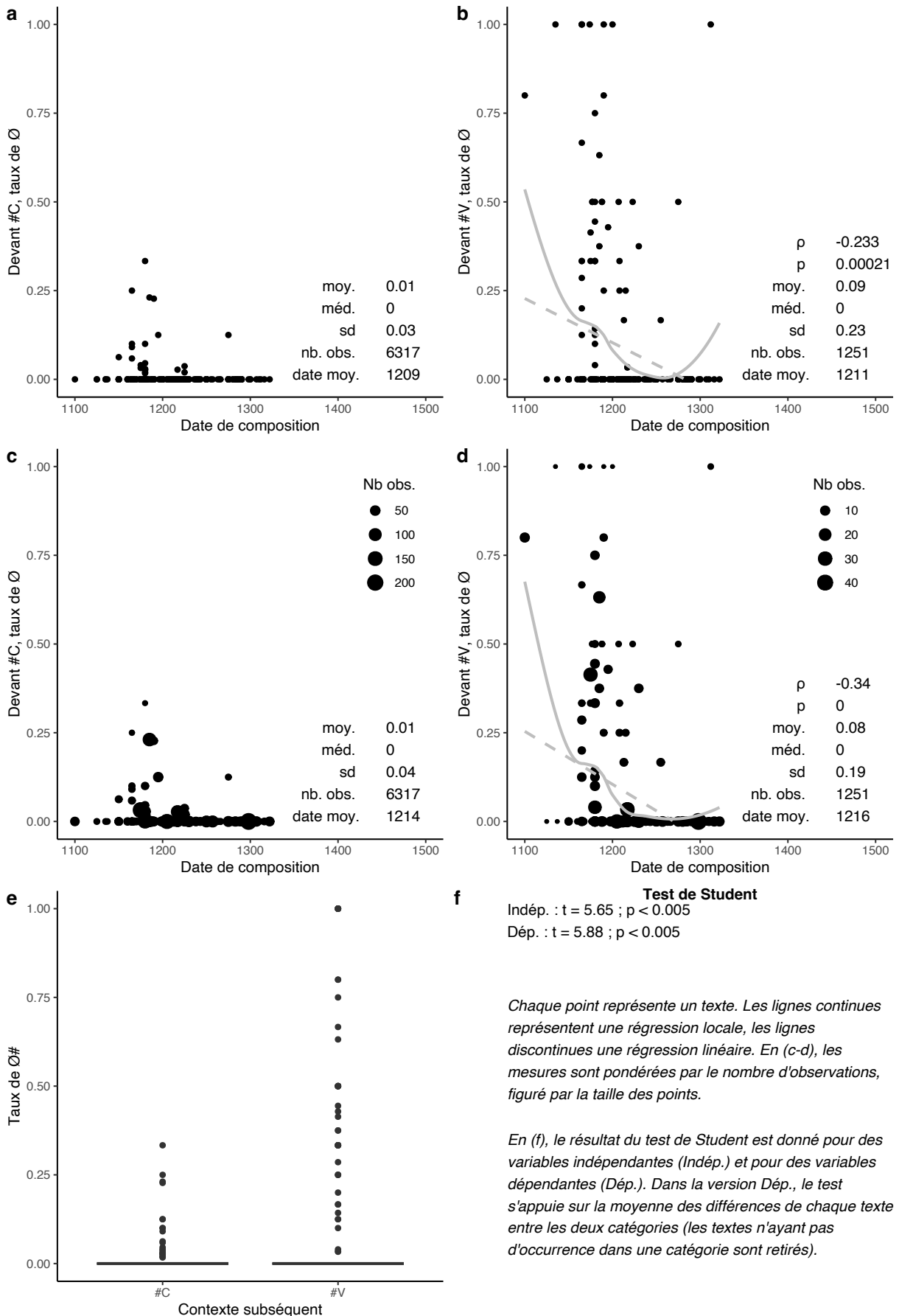


FIGURE 8.6 – Distribution contextuelle des variantes de articles indéfinis féminins singuliers en défaut de <e>.

- (8.26) *d un amur dunt il se uanta*
« d'un amour dont il se venta »
(lanvalH, p. 9, trad. pers.)
- (8.27) *le precius martir edmund de vn autre miracle enbeli*
« le précieux martyr Edmund, auréolé d'un autre miracle »
(edmund, p. 53, trad. pers.)
- (8.28) *en un tens avint jadis un autre foiz que li deable se rsemblerent*
« il advint jadis, à une autre occasion, que les démons se rassemblèrent »
(anth, p. 18, trad. pers.)
- (8.29) *vos mosterai un aventure*
« je vous montrerai une aventure »
(guigh, p. 1, trad. pers.)

Parmi les occurrences de *un* qui sont bien féminines, un nombre important se trouve devant l'adjectif *autre*, interjeté entre l'article et le nom déterminé (8.28); ces constructions *un + autre + nom* représentent 41 occurrences sur les 102 occurrences de *un* féminin devant initiale vocalique. Il semble donc, ici, que la construction syntaxique « une + autre » soit en voie de grammaticalisation, laquelle permet au <e> de chuter par l'action conjointe de la disparition de la frontière de mot (BYBEE 2006) et de la réduction du contenu segmental des constructions tokénisées (BYBEE 2000, 2002, 2006; BYBEE et SCHEIBMAN 1999).

8.2.3.2 Défauts devant #C

L MascCSsg. Seules quatre occurrences *l* pour l'article défini masculin CSsg se trouvent devant consonne, sur 24 940 occurrences dans ce contexte. Sur ces quatre occurrences, deux représentent une situation d'enclise classique, après *que* (8.30) et *ne* (8.31). Avec seulement deux occurrences non enclitiques et deux occurrences enclitiques pour 24 940 occurrences préconsonantiques, on peut considérer que le défaut vocalique de *li* devant initiale consonnantique est exceptionnel, la règle étant la préservation de la voyelle dans ce contexte.

- (8.30) *car trop fu la parole dure que l seneschaus le fist entendre*
« car le discours que tint le sénéchal fut si dur »
(jouf, p. 5, trad. pers.)
- (8.31) *ne l fiz a sun pere ne frere a sa sorur*
« ni le fils envers son père, ni le frère envers sa sœur »
(sergbH, p. 9, trad. pers.)

Un Fémsg. Devant initiale consonnantique, Comparalem a détecté 74 occurrences *un* annotées comme étant des articles indéfinis féminins singuliers. Sur ces 74 occurrences, un certain nombre sont issues d'une erreur d'annotation en genre, notamment 25 occurrences *un poi* (« un peu »). Nous n'avons trouvé que 20 occurrences dans lesquelles *un* représente assurément l'article indéfini féminin; quatre exemples en sont donnés en (8.32) - (8.35).

- (8.32) *un damaisele ismayne ot non*
« une demoiselle qui s'appelait Ismène »
(ipo, p. 17, trad. pers.)

lanvalH : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. H; Dees : nil; compo. N.-O., ca 1165; ms. a.n., ca 1275; MS2 (MARIE DE FRANCE 1958); DEAF : MarieLanvR².

edmund : DENIS PIRAMUS (1195*). *La Vie Seint Edmund le Rei*. Dees : [Angleterre]; compo. a.n., ca 1195; ms. a.n., ca 1325; MS2 (DENIS PIRAMUS 1935); DEAF : EdmK.

anth : WAUCHIER DE DENAIN (1208*). *La vie de saint Antoine*. Dees : Haute-Marne; compo. wall., ca 1208; ms. pic., ca 1275; MS1 (transcr. inédite Th. Venckeleer)

guigh : MARIE DE FRANCE (1165*). *Le Lai de Guigemar*. Ms. H; Dees : nil; compo. N.-O., ca 1165; ms. a.n., ca 1275; CR (transcr. inédite sans transcripteur indiqué)

jouf : ANON. (1250*). *Joufroi de Poitiers*. Dees : Haute-Marne; compo. SE, ca 1250; ms. Est., ca 1300; MS3 (FAY et GRIGSBY 1972); DEAF : JoufrF.

sergbH : GUISCHART DE BEAULIU (ca 1190). *Sermon*. Compo. a.n., ca 1190; ms. a.n., ca 1250; CR1 (BEAULIU 1909), DEAF : GuischartG.

ipo : HUE DE ROTELANDE (ca 1185). *Ipomedon*. Compo. a.n., ca 1185; ms. a.n., ca 1190; CR (transcription inédite).

- (8.33) *si meine un dameisele a mei semble q ele est mut bele*
 « et il conduit une demoiselle dont il me semble qu'elle est fort belle »
 (ipo, p. 57, trad. pers.)
- (8.34) *mes un chose uus dirai*
 « mais je vous raconterai quelque chose »
 (lanvalH, p. 4, trad. pers.)
- (8.35) *mis sa poinne un peau d eigneaus que il ot*
 « il mis sa broche, une peau d'agneau qu'il avait » [en énumération]
 (martin3, p. 25, trad. pers.)

lanvalH : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. H; Dees : nil; compo. N.-O., ca 1165; ms. a.n., ca 1275; MS2 (MARIE DE FRANCE 1958); DEAF : [MarieLanvR²](#)

martin3 : PÉAN GÂTINEAU (1125*). *Vie de saint Martin*. Dees : Indre, Cher; compo. tourr., ca 1125; ms. berr., ca 1290; MS (PÉAN GÂTINEAU 1896); DEAF : [PeanGatS1](#)

Il convient donc de réviser le taux de 1% de défauts devant initiale consonantique; si l'on ne conserve que 20 occurrences *un* pour *une* devant initiale consonantique, le taux de défauts chute à 0,03%.

8.2.3.3 Distribution diatopique

Comme illustré dans les deux cartes *infra*, les défauts vocaliques de *li* CSsg et *une* suivent une distribution diatopique remarquable, les défauts étant principalement concentrés à l'Ouest du domaine. Dans ce qui suit, nous détaillons chacune de ces distributions.

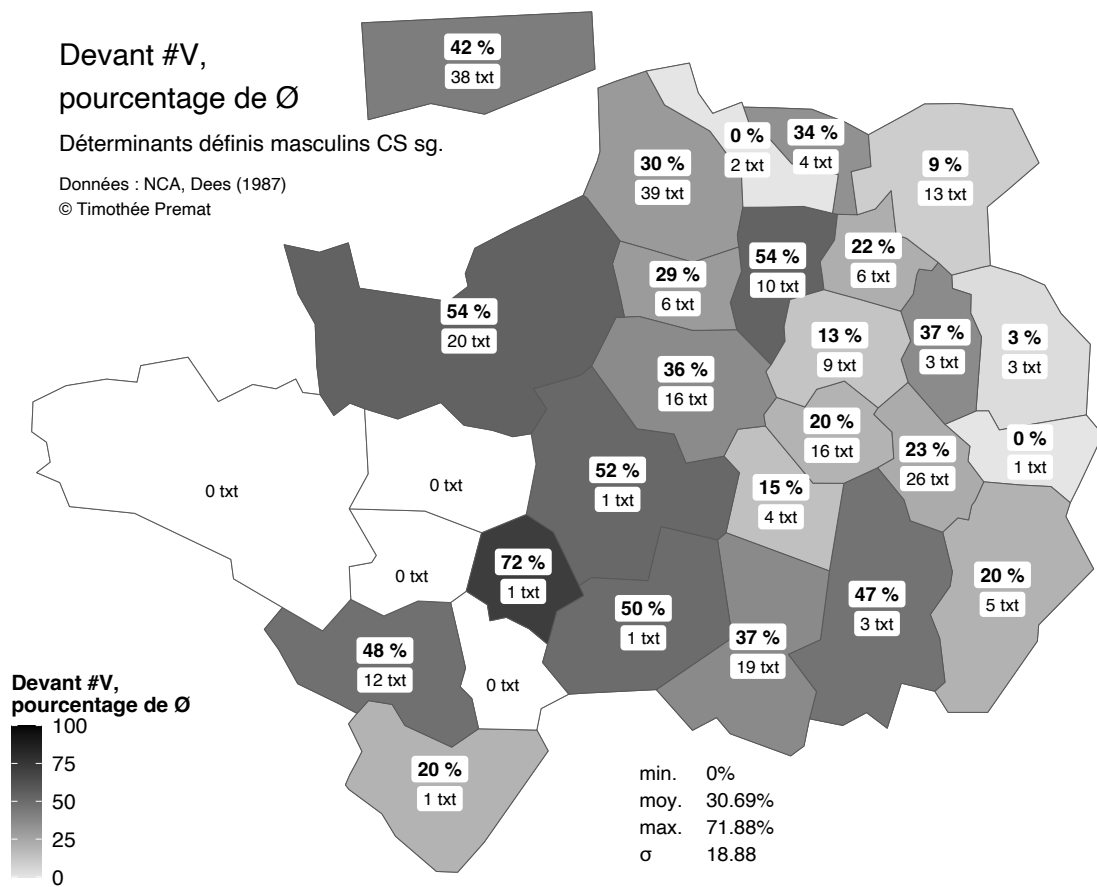


FIGURE 8.7 – Distribution diatopique des alternances <V>/Ø, sur l'ensemble de la période considérée, articles définis masculins singuliers au cas sujet (*li*) suivis d'une initiale vocalique.

L MascCSsg. Comme indiqué sur la carte en fig. 8.7, l'élision de *li* en *l* devant initiale vocalique semble particulièrement productive à l'Ouest et au Sud-Ouest du domaine, tandis que l'Ouest et particulièrement le long des marges Nord-Ouest du domaine semble préférer la préservation de *li* devant initiale vocalique. Il en résulte une situation relativement hétérogène, comme l'indique l'écart-type des taux régionaux moyens ($\sigma = 19$). Si la valeur extrême de 72% dans l'Indre-et-Loire doit être nuancée puisqu'elle ne porte que sur un seul texte, il convient de noter que d'autres taux occidentaux ou sud-occidentaux élevés sont consolidés par un nombre de textes satisfaisant (Angleterre : 42%, 38 textes ; Normandie : 54%, 20 textes ; région Nièvre, Allier : 37%, 19 textes ; région Vendée, Deux-Sèvres : 48%, 12 textes).

Au sein de cet ensemble occidental, le mitage est faible : seule la Charente-Maritime dispose d'un taux inférieur à la moyenne, et celui-ci ne porte que sur un texte. En revanche, à l'Est et au Nord-Est, le mitage est plus important, avec des régions dont les taux sont supérieurs à la moyenne (Aisne : 54%, 10 textes ; Hainaut : 34%, 4 textes ; Meuse : 37%, 3 textes) et d'autres dont les taux sont inférieurs à la moyenne voire nuls (Nord : 0%,

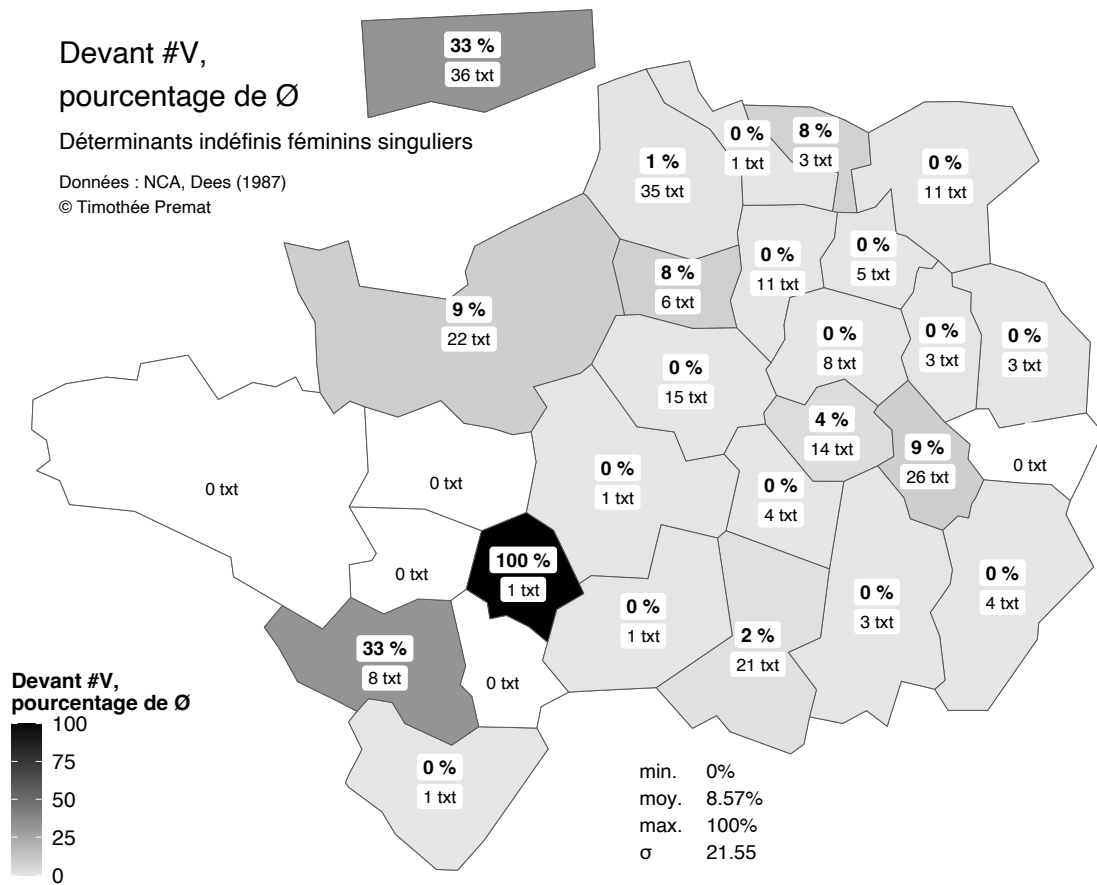


FIGURE 8.8 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, articles indéfinis féminins singuliers (*une*) suivis d'une initiale vocalique.

2 textes ; Vosges : 0%, 1 texte ; région Moselle, Meurthe-et-Moselle : 3%, 3 textes ; Marne : 13%, 9 textes, etc.). Si cette distribution et notamment son mitage à l'est et au Nord-Est est fluctuant dans le temps (cf. cartes en fig. E.18 et E.19, rejetées en annexes p. 627 et 628), c'est principalement parce que le découpage en demi-siècles divise le nombre de textes sur chaque carte et rend les taux difficilement interprétables. La situation d'ensemble demeure assez proche de celle de la carte diachronique en 8.7, et il faut donc convenir que l'élosion de l'article masculin CSsg *li* est bien plus pratiquée à l'Ouest et au Sud-Ouest du domaine qu'ailleurs ; à notre connaissance, ce fait n'avait pas encore été documenté.

Un FÉmsg. Si les occurrences défectueuses *l* pour *li* sont concentrées à l'Ouest et au Sud-Ouest, les occurrences *un* pour *une* devant initiale vocalique sont associées plus spécifiquement au dialecte anglo-normand et à la région Vendée, Deux-Sèvres, comme illustré en fig. 8.8. Ici, comme pour le défini *li*, le taux anormalement élevé de la région Indre-et-Loire est à nuancer puisqu'il repose sur un seul texte. En revanche, les taux de l'Angleterre (33%, 36 textes) et de la région Vendée, Deux-Sèvres (33%,

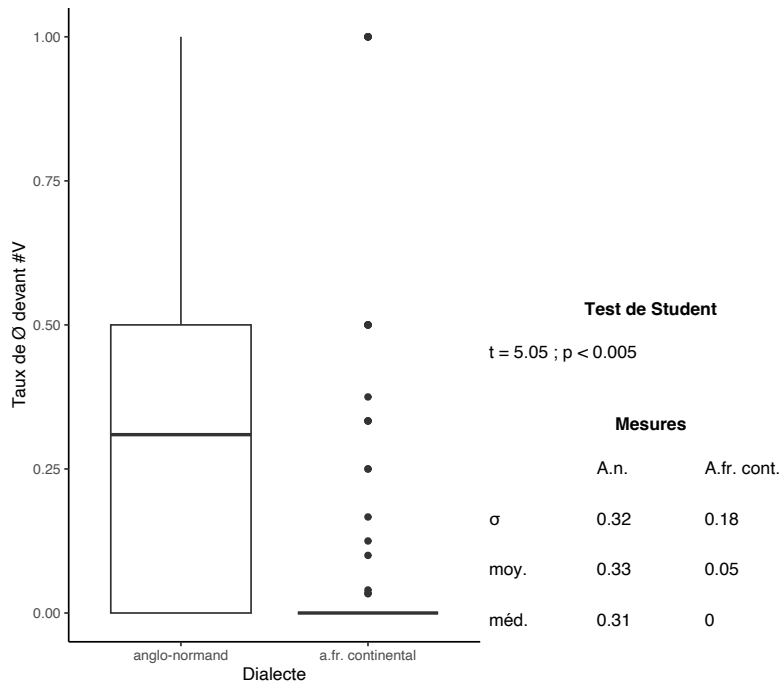


FIGURE 8.9 – Diachronie des taux de défaut de <e>, anglo-normand vs ancien français continental, articles indéfinis féminins singuliers *une* devant initiale vocalique.

8 textes) sont consolidés. Il y a donc une coïncidence partielle entre la distribution des élisions de *li* et des formes *un* pour *une* devant initiale vocalique, mais cette coïncidence n'est que partielle dans la mesure où les autres régions qui pratiquaient fortement l'élision de *li* ne pratiquent pas spécialement le défaut de <e> pour *une*.

En ce qui concerne l'anglo-normand, et comme indiqué en fig. 8.9, on retrouve ici des caractéristiques déjà relevées pour d'autres parties du discours, à savoir la plus forte hétérogénéité (a.n. : $\sigma = 0,32$; a.fr.cont. : $\sigma = 0,18$) et un taux moyen plus élevé (moy. a.n. = 33% ; moy. a.fr.cont. = 5%) en anglo-normand que sur le continent. Comme l'indiquent les médianes, la moitié des textes anglo-normands a un taux de défaut compris entre 31 et 100% de défauts, tandis que sur le continent, la médiane de 0 et l'absence d'étendue du troisième quartile au-dessus de la médiane indiquent que 75% des textes n'ont aucun défaut de <e> pour *une*. De ce fait, la tendance à la baisse indiquée par les régressions linéaires et locales en fig. 8.6.b,d n'est qu'un effet de la diatopie, dû à la précocité des textes anglo-normands. Et, comme suggéré par la carte en fig. 8.8, la moitié des textes continentaux disposant de taux positifs de défauts pour *une* correspondent aux huit textes de la région Vendée, Deux-Sèvres, et le texte à n'avoir qu'*un* à la place d'*une* est le texte d'Indre-et-Loire.

Ces données ne sont certes pas consolidées par l'épurement des occurrences défectueuses, dont nous avons relevé *supra* qu'elles n'étaient pas toutes fiables, mais une vérification manuelle partielle nous permet d'estimer que la prévalence de l'anglo-normand pour cas d'élision de *une* demeure pertinente. Enfin, pour les cas de *un* pour *une* devant initiale consonantique, l'élimination de la plupart des occurrences proposée

supra confirme une certaine prévalence de l'anglo-normand, mais le faible nombre d'occurrences (20 occ. défectueuses) ne permet guère de conclusion catégorique.

8.2.3.4 En somme

L'article défini masculin singulier au cas sujet *li* connaît, sur l'ensemble de la période, des formes élidées et des formes non élidées (devant initiale vocalique), et, au-delà du phénomène d'enclise, quelques occurrences accidentelles et négligeables d'apocope devant initiale consonantique. Si, devant initiale vocalique, on trouve en moyenne plus de formes *li* que de formes *l*, nous avons vu que la métrique pouvait indiquer que certains *li* sont à élider pour respecter l'isosyllabisme du vers, et que l'élosion graphique ne représente donc qu'une partie des élisions métriques. Quant à l'article indéfini féminin *une*, les occurrences relevées par Comparalem et représentées sur les graphiques sont sur-estimées, n'ayant pas été préalablement nettoyées d'un certain nombre d'erreurs d'annotation. L'élimination des formes suspectes ou fausses conduit à réduire de 1% à 0,3% le nombre d'occurrence défectueuses *un* devant initiale consonantique; devant initiale vocalique, s'il faut également retirer certaines occurrences, les attestations sont plus solides, et concernent notamment 41 constructions *un + autre + nom* où le nom déterminé est féminin, et concernent principalement le domaine anglo-normand et la région Vendée, Deux-Sèvres. De ce fait, *une* semble ici se comporter plutôt comme certains adjectifs que comme les autres articles à défauts vocaliques, c'est-à-dire qu'il est susceptible de perdre marginalement son <e> final devant initiale vocalique, particulièrement dans des constructions grammaticalisées, mais son élosion n'a pas du tout le caractère de systématisme de l'élosion de *le* ou *la*, ni le degré d'optionnalité de celle de *li*.

8.2.4 Article à élosion prohibée : *li* (CSpl)

Comme indiqué en table 8.1 et en fig. 8.10 en page suivante, l'article défini de CS pluriel masculin *li* ne connaît pas de défauts significatifs quel que soit le contexte. Comme indiqué par la version pondérée du graphique de série temporelle du contexte prévocalique en fig. 8.10.d, les cinq textes comportant des occurrences *l* annotées comme représentant l'article *li* sont dans des textes qui disposent de peu d'occurrences de celui-ci, et aucun texte ne dispose de plus de deux occurrences *l* pour *li*. De fait, avec 8 occurrences *l* pour *li* (devant #C comme devant #V), les occurrences de défaut vocalique de *li* sont si marginales que l'on peut se passer de les examiner, en s'appuyant sur le taux pondéré devant initiale consonantique (< 1%).

Il convient de noter que *li* pluriel, pourtant constitué du même matériel phonologique que *li* singulier, exprimant le même cas et se trouvant *a priori* dans les mêmes contextes syntaxiques, ne connaît pas l'élosion optionnelle que connaît *li* singulier. Cet élément, capital pour l'analyse que nous proposerons en chap. 14, signifie que l'homographie des articles définis CS sg et pl ne veut pas nécessairement dire qu'ils ont la même représentation phonologique.

Déterminants définis masculins pluriels CS

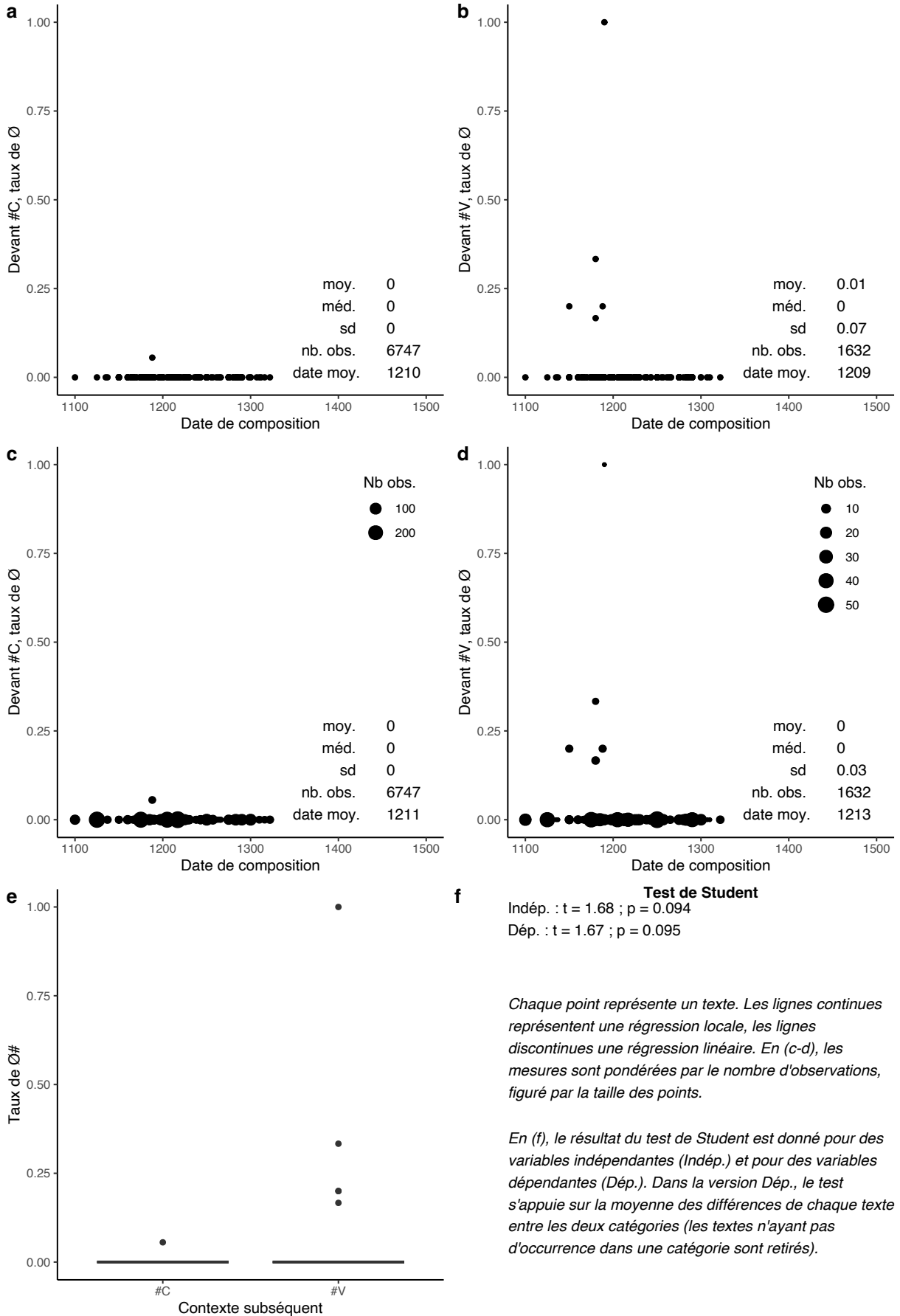


FIGURE 8.10 – Distribution contextuelle des variantes de articles définis masculins pluriels au cas sujet en défaut de <i>.</i>

8.3 Des observations aux généralisations

Dans cette section, nous regroupons les observations faites sur les données ci-dessus pour obtenir des généralisations aptes à supporter un discours analytique. Nous commençons par reprendre la question de la distinction entre les articles qui sont atteints par l'élision et ceux qui ne le sont pas (sec. 8.3.1), avant de reprendre sous une forme synthétique les observations faites sur le rôle du contexte, qui jusqu'ici ont principalement été énoncées sous forme d'opposition entre ce qui tient de l'élision et ce qui n'en tient pas (sec. 8.3.2). Nous terminons par reprendre les quelques remarques diatopiques et diachroniques formulées ci-dessus (sec. 8.3.3).

8.3.1 Lemmes atteints par l'enclise et l'élision

Tous les articles ne sont pas atteints par les défauts de voyelle finale, et ne le sont pas de manière univoque. Néanmoins, au sein de ceux qui *pourraient* être atteints par un défaut vocalique, tous sauf *li* MASC CSpl sont atteints, à des degrés divers, par un tel défaut, principalement devant initiale vocalique (cf. sec. 8.3.2). Il y a donc une forme de sélectivité qui doit s'appliquer pour déterminer quels articles seront sujets à l'élision, obligatoire ou facultative, et lesquels y échapperont. Cette sélectivité ne saurait être syntaxique, dans la mesure où les déterminants féminins singuliers s'élident au CS alors que les déterminants masculins ne s'y élident que facultativement au singulier et résistent à l'élision au pluriel; or la syntaxe des déterminants au cas sujet est essentiellement la même, peu importe leur genre et leur cas. La sélectivité est également difficile à mettre en place sur un critère morphologique, dans la mesure où elle demanderait de séparer le CS masculin singulier (*li* : élision facultative) et le CS féminin singulier (*la* : élision systématique), et où le CS masculin pluriel n'a pas d'équivalent élidable au féminin (*elles* : la présence de *-s* bloque l'élision). Ces remarques sont formalisées en (8.36) et (8.37).

(8.36) **ART : sélectivité lexicale.**

Il est impossible, sur la base de critères phonologiques, de prédire quels articles sont atteints par l'élision.

(8.37) **ART : sélectivité morpho-syntaxique.**

Ni la morphologie ni la syntaxe ne sont en mesure de prédire quels articles sont sujets à l'élisions.

Il faut donc admettre, par défaut, le principe suivant : les articles monosyllabiques à finale vocalique connaissent une élision systématique (MASC CRsg *le*, FÉM sg *la*, *le*), et ce sont les formes *li* qui se démarquent de ce paradigme. Pour ces formes *li*, il faut admettre, malgré leur homographie, que leur représentation phonologique doit différer, puisque la syntaxe ne saurait faire la différence entre le CS singulier et le CS pluriel. Si nous en venons à proposer une différence dans la représentation phonologique entre *li* sg et *li* pl, cette différence est une déduction, et ne permet en rien la prédiction de leur comportement, ce qui ne vient pas en contradiction avec (8.36).

Enfin, l'indéfini féminin singulier *une* se trouve dans une situation qui rappelle celle du pronom personnel *elle*, étant le seul membre du paradigme à être polysyllabique (cf. sec. 9.3.1, généralisation 9.60). De fait, *une* ne semble pas se comporter comme les autres articles, mais à la manière des adjectifs, ne perdant sa voyelle que rarement, et souvent dans le cadre d'une construction lexicalisée. Ces deux éléments, dont le premier semble dépendre au moins en partie du second, sont synthétisés en (8.38) et (8.39).

(8.38) **ART : restrictivité phonologique.**

Si la seule phonologie ne peut prédire quels seront articles atteints par l'élision, elle semble en revanche fournir un critère de différenciation de l'article indéfini féminin singulier, qui est le seul à être bisyllabique (*une*). Cet article se comporte comme les adjectifs féminins singuliers (cf. 8.39).

(8.39) **ARTINDÉFÉM sg : grammaticalisation.**

Environ la moitié des occurrences *un* pour *une* __#V appartiennent à la construction *un(e) + autre + [nom]^{FÉM}*, qui semble donc se grammaticaliser.

8.3.2 Le rôle du contexte

Les cas d'enclise, du moins encodés sous la forme de deux unités graphiques dans le corpus, sont marginaux pour les articles, vis-à-vis de ce que nous avons pu noter pour les pronoms. Le rôle du contexte est donc toujours unidirectionnel : lorsque les défauts sont attestés en quantité significative, c'est toujours devant initiale vocalique. Cet effet varie, selon que l'élision est systématique (MASC CR *le*, FÉM CS-R *la, le*) ou optionnelle (MASC CSsg *li*). Le contexte prévocalique est donc une condition nécessaire aux défauts de voyelle finale pour les articles définis, et cette condition est aussi suffisante pour les articles masculins singuliers CR *le* et féminins singuliers CS-R *la, le*. Ces éléments sont repris en (8.40) et (8.41). Ici, comme en (8.42) *infra*, les mesures indiquées proviennent des figures 8.1, 8.3, 8.4 et 8.6, et les valeurs-*t* et *p* sont celles du test de Student à variables indépendantes.

(8.40) **ARTDÉF : contexte, *le, la***

Le contexte prévocalique agit comme condition nécessaire et presque suffisante au déclenchement du défaut des articles définis masculins singuliers CR *le* et féminins singuliers CS-R *la, le*.

(MASC CR sg : $t = 350$; $p < 0,005$; moy. __#V = 99%)

(FÉM CS-R sg : $t = 532$; $p < 0,005$; moy. __#V = 99%)

(8.41) **ARTDÉF : contexte, *li* CSsg**

Le contexte prévocalique agit comme condition nécessaire mais non suffisante au déclenchement du défaut des articles définis masculins singuliers, dans la mesure où cette élision n'est pas systématique.

($t = 17$; $p < 0,005$; moy. __#V = 33%)

Quant à l'article indéfini féminin singulier *une*, le contexte ne joue ici qu'un plus faible rôle, dans la mesure où les défauts restent rares; la force de l'effet du contexte prévocalique est à peu près équivalente à celle trouvée pour les adjectifs féminins singuliers. Cet élément est repris en (8.42).

(8.42) **ARTINDÉF : contexte, *une* CS-Rsg**

Le contexte prévocalique agit comme condition favorable, mais ni nécessaire ni suffisante, au déclenchement du défaut des articles indéfinis féminins singuliers *une*.

($t = 6$; $p < 0,005$)

8.3.3 Diachronie et diatopie

L'effet du temps et de l'espace est nul ou négligeable pour la plupart des articles interrogés, étant donné que leur élision est systématique ou prohibée pour trois d'entre eux. En revanche, nous avons noté en sec. 8.2.3.3 que ceux pour lesquels l'élision existait sans être systématique montraient une distribution diatopique marquée en faveur de l'Ouest du domaine. En effet, l'article défini masculin singulier CS *li* s'élide plus souvent sur une large moitié Ouest et Sud-Ouest du domaine, tandis que l'article indéfini *une* perd sa voyelle finale principalement en Angleterre et en Vendée, Deux-Sèvres. Ces éléments sont repris en (8.43).

(8.43) **ART : diachronie et diatopie.**

L'ARTDÉFMASC CSsg *li* s'élide plus à l'Ouest et au Sud-Ouest qu'ailleurs; l'ARTINDÉFFÉM CS-Rsg *une* perd son *-e* principalement en Angleterre et en Vendée, Deux-Sèvres.

8.4 Conclusion

Dans ce chapitre, nous avons étudié la distribution des défauts vocaliques des articles se terminant par une voyelle. Dans l'introduction (sec. 8.1), nous avons présenté le paradigme des articles en a.fr. (sec. 8.1.1), avant d'aborder rapidement la chaîne de traitement des données mise en place (sec. 8.1.2). Nous avons ensuite présenté les résultats généraux (sec. 8.2), avant d'esquisser la typologie qui guide la suite de l'exposition (sec. 8.2.1). Nous avons ainsi mis en évidence que les articles définis masculins singuliers au cas régime *le* et féminins singuliers *la* (et la variante dialectale *le*) étaient sujets à une élision systématique (sec. 8.2.2), tandis que ce défaut n'était pas systématique pour l'article défini masculin CS *li* et l'indéfini féminin *une* (sec. 8.2.3). Quant à l'article défini CSpl masculin *li*, nous avons vu que nos données confirmaient le fait qu'il ne connaissait pas d'élision (sec. 8.2.4).

Nous avons ensuite pu résumer ces observations quantitatives à un ensemble de généralisations analytiques (sec. 8.3). Nous avons d'abord noté (sec. 8.3.1) que nous n'étions pas en mesure de prédire l'élision des articles à partir de critères phonologiques (8.36) ou morphosyntaxiques (8.37), mais que néanmoins la phonologie permettait de séparer l'indéfini féminin singulier *une* des autres articles étudiés, celui-ci étant bisyllabique (8.38), et répondant par ailleurs fortement à un effet de construction

grammaticalisées, à la manière des adjectifs féminins singuliers (8.39). L'élision, systématique, optionnelle ou prohibée, est reflétée dans nos analyses par la force de l'effet du contexte prévocalique sur la génération de défauts vocaliques, force qui est à un niveau jamais atteint jusqu'à présent pour les articles définis singuliers *le* (MAscCS) et *la* (FÉMCS-R : 8.40) mais qui est bien moins marquée pour l'article défini singulier CS *li* (8.41) et pour l'indéfini féminin singulier *une* (8.42). Enfin, nous avons noté que seuls les défauts de l'article défini pluriel *li* et de l'indéfini féminin *une* suivaient une distribution diatopique remarquable, les occurrences défectueuses étant concentrées à l'Ouest et au Sud-Ouest pour *li* et en Angleterre et en Vendée, Deux-Sèvres pour *une*, ce qui produit également un artéfact diachronique, particulièrement pour *une* (8.43). Pour les autres articles, les variables diatopiques et diachroniques semblent sans effet, ce qui est dû au caractère systématique ou prohibé des défauts de voyelle finale.

Ainsi, les articles forment une classe au comportement complexe, allant de l'élision systématique à la prohibition de l'élision. Comme pour les pronoms, les formes bisyllabiques féminines semblent sortir du paradigme général, rejoignant la variation d'une catégorie lexicale ; ici *une* se comporte à la manière de certains adjectifs féminins singuliers. Dans tous les cas, et s'il sera nécessaire de proposer une différence dans la représentation phonologique de *li* CSsg et *li* CSpl, aucun critère phonologique, morphologique ciblant une classe discrète, ou tenant à la structure syntaxique, ne permet de distinguer les formes *le* et *la* à élision systématique, ou les formes *li* de CSsg et les formes *li* de CSpl. L'information relative au caractère systématique, optionnel ou prohibé de l'élision doit donc se trouver encodée ailleurs dans la grammaire et dans ses représentations, ce sur quoi nous reviendrons dans l'analyse en chap. 14.

9.1 Introduction

Ce chapitre est dédié aux défauts de voyelle finale atone des pronoms personnels. En effet, contrairement aux parties du discours vues précédemment, les pronoms personnels connaissent une variation qui n'est pas uniquement due à la perte d'un <e> final mais peut aussi concerner d'autres voyelles, comme c'est encore le cas en fr.mod. (*tu aimes?* ~ *t'aimes?*; **il te aime* → *il t'aime*).

Étant donnée la nature de la catégorie étudiée, l'exclusion de lemmes et de formes ainsi que l'opération de relemmatisation se sont faites de manière plus radicale, aussi la présentation de ces éléments change quelque peu vis-à-vis des chapitres précédents. Après une brève introduction concernant les pronoms personnels en a.fr. (sec. 9.1.1), nous couvrons donc cette étape du traitement des données (sec. 9.1.2), avant d'aborder les résultats de nos analyses (sec. 9.2) et d'en extraire l'essentiel sous la forme de généralisations (sec. 9.3).

9.1.1 Les pronoms personnels de l'ancien français

Les pronoms personnels de l'a.fr. sont classiquement divisés en *formes fortes* (ou toniques) et *formes faibles* (ou atones), les premières étant parfois diphtonguées (*mei, tei*, etc.) et ne connaissant ni élision ni enclise, les secondes étant pour la plupart non diphtonguées (*me, te*) et sujettes à l'élision et à l'enclise (*me+en* → *m'en*; *si+me* → *sim*). L'ensemble du paradigme des pronoms personnels est reproduit en table 9.1. Deux personnes échappent tout à fait à l'opposition entre formes toniques et atones : les première et deuxième personnes du pluriel ne connaissent ni opposition entre formes fortes et formes faibles, ni opposition entre cas sujet et cas régime. Pour les première et deuxième personnes du singulier, les formes toniques et atones au cas sujet ne se différencient qu'à partir du XII^e s. : dans les plus anciens textes, il n'existe pas de distinction graphique entre une forme faible et une forme forte ; ainsi, la même forme est utilisée lorsque le pronom précède directement le groupe verbal (9.1) et lorsqu'il est détaché de celui-ci (9.2).

- (9.1) *que jo lur ai guerpide*
 « [mon épouse] que j'ai abandonnée »
 (*Alexis*, v. 208, cité par MARCHELLO-NIZIA 2020c, GGHF, p. 677, trad. pers.)
- (9.2) *E jo, dolente, cum par gui avoglie!*¹
 « Et moi, souffrante, comme je fus aveuglée! »
 (*Alexis*, v. 434, *id.*, trad. *id.*)

AlexisRaM : ANON. (1050*). *Vie de saint Alexis*. Compo. ca 1050, ms. ca 1120, dial. norm. (RAINSFORD et MARCHELLO-NIZIA 2018).

1: L'édition d'*AlexisRaM* n'a pas de virgule après *jo*.

TABLE 9.1 – Paradigme des pronoms personnels. Adapté de MARCHELLO-NIZIA (2020c, GGHF, p. 676-677).

Tonicité = '—' : les formes peuvent être toniques et atones.

Datation = '—' : les formes ne sont pas associées à une époque précise.

→ x : jusqu'au siècle x ; x → : à partir du siècle x.

« Dir. » : (cas régime en construction) directe ; « indir. » : (cas régime en construction) indirecte.

Personne	Cas	Tonicité	Datation	Formes
1sg	CS	—	→ m. XII ^e	<i>je, jo, jeo, jou</i>
		Atone	m. XII ^e →	<i>jé, ge, g, j</i>
		Tonique	m. XII ^e →	<i>je, ge, gié, jou</i>
	CR	Atone	—	<i>me, m</i>
		Tonique	—	<i>mei, moi, mai, mi (pic.)</i>
2sg	CS	—	→ XII-XIII ^e	<i>tu, te</i>
		Atone	XII ^e → XIV-XV ^e	<i>tu, te, t</i>
		Tonique	XII ^e → XIV-XV ^e	<i>tu</i>
	CR	Atone	—	<i>te, t</i>
		Tonique	—	<i>tei, toi, ti (pic.)</i>
MAsc3sg	CS	—	→ XII ^e	<i>il, i</i>
		Atone	XII ^e → XIV-XV ^e	<i>il, i</i>
		Tonique	XII ^e → XIV-XV ^e	<i>il</i>
	CR	Atone	-	<i>le, l (dir.); li, en, i (indir.)</i>
		Tonique	-	<i>lui</i>
FÉM3sg	CS	—	-	<i>ele, el, al (Ouest)</i>
	CR	Atone	-	<i>la, l, le (pic.) (dir.); li, en, i (indir.)</i>
		Tonique	-	<i>li, lié (Ouest), lei (Est) (dir.); li (indir.)</i>
NEUT3sg	CS	Atone	→ XV ^e	<i>il, el, al, eil</i>
	CR	Atone	-	<i>le, lo, l (dir.); en, i (indir.)</i>
1pl	—	—	-	<i>nos, nus (a.n.), nous</i>
2pl	—	—	-	<i>vos, vus (a.n.), vous</i>
MAsc3pl	CS	—	-	<i>il</i>
	CR	Atone	-	<i>les, s (dir.); lor, leur (indir.)</i>
		Tonique	-	<i>els, eus, aus, ax, iaus</i>
FÉM3pl	CS	—	-	<i>eles, els, eus (Ouest), als (Ouest)</i>
	CR	Atone	-	<i>les (dir.); lor, leur (indir.)</i>
		Tonique	-	<i>eles, els, eus (Ouest)</i>
RÉFL.3sg-pl	CR	Atone	-	<i>se, s</i>
		Tonique	-	<i>sei, soi, si (pic.), sai</i>

9.1.1.1 Je CS : progression de l'élision

Parallèlement à l'absence de distinction entre formes fortes et faibles, les pronoms sujet de première et deuxième personne du singulier ne s'élident pas avant le XII^e s. Dans son étude dédiée à l'élision du pronom « je », MARCHELLO-NIZIA (2015) ne trouve aucune occurrence où le pronom sujet se réduit à sa forme consonantique et/ou où la métrique indiquerait une élision non transcrite. En revanche, dans *Roland*, si *jo* n'est jamais écrit *j*, un certain nombre de vers appellent à l'élision de *jo* devant initiale vocalique pour que le vers soit bien formé; MARCHELLO-NIZIA (2015, p. 27) en compte 5 (9.3 - 9.6),² et, en utilisant le PAM pour détecter les vers hypermétriques (PREMAT et POGGIO [à paraître]), nous en comptons 9,

MARCHELLO-NIZIA (2015). « De JE à J' en français : étapes vers l'élision, interactions entre phonétique et syntaxe ».

2: Dans son article, l'auteur parle de cinq exemples mais n'en donne que quatre. Nous ne savons lequel de nos 5 exemples supplémentaires correspondrait à celui qu'elle ne cite pas.

PREMAT et al. ([à paraître]). « Restituer des fragments de phonologie avec le PAM (*Programme d'Analyse Métrique*) : l'élision de schwa et des voyelles pleines dans les monosyllabes fonctionnels en très ancien français ».

en ajoutant ceux en (9.7 - 9.11). Dans ces exemples, sont placées entre chevrons les syllabes extramétriques.

- (9.3) *Respunt dux Nei(mes)* : « *Jo irai, par vostre dun !*³
« Le duc Naime répond : « J'irai, si vous me l'accordez »
(*Roland*, v. 246, trad. MARCHELLO-NIZIA 2015, p. 26)
- (9.4) *Respunt Rollant* : « *Jo i puis aler mult ben !*⁴
« Roland répond : « Je peux très bien y aller »
(*Roland*, v. 254, trad. *id.* p. 27)
- (9.5) *Jo i puis aler, mais n'i avrai guarant*⁵
« Je peux y aller, sans avoir aucun garant »
(*Roland*, v. 329, trad. *id.*)
- (9.6) *Se jo ai parenz, n'en i ad nul si proz.* »⁶
« Si j'ai des parents, il n'y en a aucun d'aussi valeureux »
(*Roland*, v. 2905, trad. *id.*)
- (9.7) *Ive e Ivo(rie), que jo aveie tant chers ?*⁷
« Ivon, Ivoire, que j'aimais tant ? »⁸
(*Roland*, v. 2406, trad. SHORT 1990)
- (9.8) *Li .XII. per, que jo aveie laiset ?* »⁹
« Et les douze Pairs que j'avais laissés ici »
(*Roland*, v. 2410, trad. *id.*)
- (9.9) *U est Marsi(lie) que jo aveie mandet ?* »¹⁰
« Où est Marsile, que j'avais convoqué ? »
(*Roland*, v. 2770, trad. *id.*)
- (9.10) *Livrez le mei, jo en ferai la justi(ce.)* »¹¹
« Livrez-le-moi, j'en ferai justice ! » »
(*Roland*, v. 498, trad. *id.*)
- (9.11) *Jo oi al corner que guaires ne vivrat.*¹²
« au son du cor j'entends qu'il ne vivra plus guère. »
(*Roland*, v. 2108, trad. *id.*)

3: SHORT (1990) édite directement : *j'irai*.

roland: ANON. (ca 1100). *Chanson de Roland*. Compo. ca 1100, ms. ca 1120, dial. norm. (MOIGNET 1972a).

4: SHORT (1990) : *j'i*.

5: SHORT (1990) : *J'i*.

6: SHORT (1990) : *j'ai*.

7: SHORT (1990) : *j'avaie*.

8: Dans une liste de personne décédées.

9: SHORT (1990) : *j'avaie*.

10: SHORT (1990) : *j'avaie*.

11: SHORT (1990) : *j'en*.

12: SHORT (1990) : *j'oi*.

Dans ces neuf vers, si *jo* compte pour une syllabe métrique, le vers devient hypermétrique. Certains éditeurs, dont SHORT (1990), par ailleurs spécialiste de l'anglo-normand, n'hésitent pas à corriger d'emblée en appliquant l'élision du pronom. L'élision n'est donc, au début du XII^e s., pas encore indiquée au niveau graphique, mais on peut considérer qu'elle est impliquée par la métrique. Si l'implication métrique de l'élision ne dit rien de son statut dans la langue non métrique, le fait que les formes graphiques se réduisant à la seule consonne *j* (ou *i*) apparaissent dès ca 1150 rend plausible le fait que l'élision métrique de *Roland* reflète une propriété de la langue en général. La première attestation que trouve MARCHELLO-NIZIA (2015, p. 28) au sein de son corpus est dans le *Lapidaire alphabétique* attribué à Philippe de Thaon, dont la composition date du premier tiers du XII^e s. et la copie de ca 1200.

- (9.12) *Ke cet essample que j'ai donné*
Qui est prové et esprové
« Car cet exemple que j'ai donné,
Qui est prouvé et éprouvé »
(*Lapidaire alphabétique*, v. 281-282, trad. MARCHELLO-NIZIA 2015)

Lapidal : PHILIPPE DE THAON (ca 1133*). *Lapidaire alphabétique*. Compo. 1^{er} t. XII^e s., ms. ca 1200, dial. a.n. (THAON 1924).

CAZAL, PARUSSA et LLAMAS-POMBO (2020). « Graphies : des usages à la norme ».

LLAMAS-POMBO (2020). « Ponctuation ».

Ici, comme dans l'édition de SHORT (1990) de *Roland*, l'apostrophe est un signe éditorial, dont l'usage n'est inventé qu'au XVI^e s. et généralisé

au xviii^e s. (GGHF, p. 515, 602); dans les manuscrits médiévaux, ce qui est rendu par bien des éditions modernes avec une apostrophe est une agglutination des deux mots.¹³ Ainsi, dans l'exemple en (9.12), c'est *jai* ou *iai* qui est rendu par *j'ai*. MARCHELLO-NIZIA (2015, p. 28) considère cette agglutination, par ailleurs associée à une seule position métrique, comme un témoin fiable de l'élision.

Dans tous ces exemples, le contexte de l'élision est strictement défini : *je* (et *tu*) ne s'élident que devant un verbe ou un pronom. Même dans ce contexte, l'élision n'est jamais systématique : dans *Le Couronnement de Louis* (compo. ca 1130, ms. xiii^e s.), MARCHELLO-NIZIA (2015) note 7 formes non élidées (et dont le pronom correspond à sa propre position métrique) pour 17 formes élidées (et dont le pronom ne correspond pas à une position métrique). Elle illustre cette opposition en comparant *je oeie* et *j'atent* en (9.13) et (9.14).

- (9.13) *Se je oeie novele al repaïrer*
Que Looïs i eüst encombrïer
 « Si j'entendais à notre retour la nouvelle
 que Louis ait des difficultés »
 (*Le Couronnement de Louis*, v. 1990-1991, trad. MARCHELLO-NIZIA 2015)
- (9.14) *Par Mahomet, dont j'atent le pardon*
Onc mais par ome n'oi tel destrucion
 « Par Mahomet, dont j'espère le pardon,
 jamais je n'ai reçu une telle blessure de quelqu'un. »
 (*Le Couronnement de Louis*, v. 962-963, trad. *id.*)

À la fin du xii^e s. et au début du xiii^e s., l'élision n'est plus restreinte aux seuls contextes où *je* précède le verbe ou un pronom. Enfin, ce n'est qu'au milieu du xiv^e s. que l'élision de *je* devient systématique devant initiale vocalique, du moins dans certains textes (chez Machaut, par exemple), tandis que d'autres auteurs conservent des occurrences atones *je* ou *ge* non élidées, MARCHELLO-NIZIA (2015, p. 36-37) notant néanmoins une tendance de plus en plus forte à l'élision devant le verbe, qui aboutit à la fin du xv^e s. et au xvi^e s. à la situation moderne, où l'élision est obligatoire devant initiale vocalique.¹⁴

Tu (te), chemin parallèle? Dans les premiers textes, le pronom sujet de 2sg est toujours *tu* ne connaît pas l'élision. Au xii^e s., apparaît la forme *te* au CS, principalement au Nord et à l'Est; celle-ci procède peut-être de l'analogie avec *je*, et c'est à cette époque que le pronom de sujet de 2sg connaît ses premières élisions (MARCHELLO-NIZIA 2020c, GGHF, p. 679-681). Il semble donc suivre une évolution parallèle à celle du pronom sujet de première personne, avec d'abord une série unique de formes non élidables, puis l'apparition de deux séries (partiellement redondantes), avec une forme forte (*tu*) non élidable et des formes faibles (*tu*, *te*) élidables. Cette élision, ici, est restée optionnelle, n'atteignant pas le stade de la généralisation de l'élision de *je* __#V à la fin du Moyen Âge; encore aujourd'hui, l'élision de *tu* sujet est un trait oral familier (GGHF, *id.*, p. 681). Dans nos données, la similitude entre l'évolution des pronoms sujets de 1-2sg est bien plus limitée que ce que rapporte la GGHF; nous n'avons que très peu de cas d'élisions de *tu*, *te* sujet, et encore, une partie

13: S'il n'y a pas d'exemple de *scripta continua* pour l'a.fr., c'est-à-dire qu'on trouve toujours une forme de frontière entre certaines unités graphiques dans les manuscrits transcrivant l'a.fr., le mot graphique médiéval demeure irréductible au mot graphique moderne, et l'agglutination peut, selon le type d'écriture et le scribe, être aussi bien contingente que grammaticale ou rythmique (CAZAL, PARUSSA et LLAMAS-POMBO 2020, GGHF, p. 514-515).

Hors du corpus BFM : éd. de LANGLOIS (1925).

14: La systématisation de l'appartenance du pronom sujet au domaine verbal est reprise développée dans le dernier chapitre, sec. 14.3.2.2.

des cas relevés est suspecte voire fautive (cf. sec. 9.2.7 *infra*). Ainsi, si l'on se limite aux données que le corpus du NCA permet de traiter, il convient de limiter fortement le parallèle entre ces deux pronoms, la deuxième personne du singulier *tu, te* ne connaissant pas la tendance à l'élision que connaît la première personne *jo, je*.

9.1.1.2 Pronoms régimes

À la différence des pronoms sujets de 1-2sg, les pronoms régimes faibles qui se terminent par une voyelle connaissent tous l'élision, et ce dès les plus anciens textes, comme en atteste la forme *le+lint* → *lint* en (9.15). Cette élision est obligatoire dans le même contexte que celui qui déclenchait les premières élisions du pronom sujet de 1sg, c'est-à-dire devant un groupe verbal introduit par *en* ou *i* et devant le verbe lui-même, s'il est à initiale vocalique. Il n'y a que le pronom régime indirect *li* (« lui », MASC/FÉM) pour qui l'élision n'est pas automatique, là où le fr.mod. ne pratique d'ailleurs pas l'élision non plus (9.16), n'ayant conservé que la forme forte du masculin *lui* pour les deux genres (GGHF, *id.*, p. 685).

- (9.15) *non lint pois*
« je ne l'en peux » [empêcher]
(Strasbourg, cité et trad. par MARCHELLO-NIZIA 2020c, GGHF)
- (9.16) *onques ne l'en sovint*
« Il ne lui en souvint pas du tout »
(Aucassin, p. 10, cité et trad. *id.*)

strasbBfm : ANON (842*). *Serments de Strasbourg*. Compo. ca 842, ms. ca 1000, (LAVRENTIEV, RAINSFORD et GUILLOT-BARBANCE 2014).

aucassin : ANON. (1208*). *Aucassin et Nicolette*. Compo. dern. q. XII^e s. ou 1^{ère} m. XIII^e s., ms. fin XIII^e s., dial. pic. (ROQUES 1929a).

9.1.1.3 Circularité de l'élision et de l'identification des formes faibles, et nature du phénomène de défaut

Lorsqu'une forme faible est homographe d'une forme forte, les grammairiens s'accordent à considérer qu'on ne peut déterminer le statut de la forme que si elle s'élide (POPE 1966 [1934], p. 321; MOIGNET 2002 [1976], p. 53-54; ZINK 1989, p. 85; HASENORH 1993, p. 71; MARCHELLO-NIZIA 2015, p. 20). Ainsi, *je, jo, jeo, jou, ge* et *tu* sont ambigus quant à leur tonicité, tandis que *j, g* et *t* indiquent clairement que la forme est élidée. Il y a là un problème pour notre recherche, qui cherche à établir s'il existe des formes défectueuses (dont la voyelle finale manque). En effet, (a) la potentielle élision d'une forme forte la réduirait à sa seule consonne, ce qui la reclasserait immédiatement comme une forme faible, et (b) l'impossible élision des formes fortes est fondée sur le fait qu'on ne trouve pas de formes fortes élidées. Ergo, (c) l'impossible attestation de formes fortes élidées (b) découle directement de (a) et n'est pas simplement liée au principe qu'on ne peut jamais attester de l'inexistence de quelque chose, mais provient ici d'un biais de circularité. Le problème vient donc du fait qu'une décision de typologie morpho-syntaxique se fonde sur un critère grapho-phonologique, ce qui rend ce critère grapho-phonologique impossible à étudier en fonction de la typologie morpho-syntaxique. Le point (c) ne rend pas simplement impossible l'attestation de formes fortes élidées, il rend aussi impossible de savoir si les *je* non élidés devant un verbe — cf. (9.13) *supra* — sont dus au fait que l'élision n'est pas systématique ou au fait qu'il s'agit du *je* issu de la série des formes fortes et non de son homographe issu de la série des formes faibles. En somme,

ZINK (1989). *Morphologie du français médiéval*.

HASENORH (1993). *Introduction à l'ancien français*.

le caractère facultatif de l'élision est tout aussi impossible à établir que l'absence d'élision des formes fortes.

Il ne s'agit pas, pourtant, de remettre en question l'existence, bien avérée et consensuelle, des séries fortes et faibles des pronoms. Dans ce qui suit, et étant données les possibilités permises par notre corpus, nous excluons les formes explicitement fortes (dotées d'une graphie distinctive), mais conservons toutes les formes ambiguës au côté des formes faibles. De fait, nos données sont ici biaisées par l'impossibilité de distinguer entre formes faibles et formes fortes lorsqu'une forme appartient aux deux séries. Néanmoins, les résultats de nos investigations demeurent cohérents avec le chemin tracé par l'étude de MARCHELLO-NIZIA (2015), ce qui leur donne une certaine validité, malgré la confusion des deux séries.

Enfin, puisque les pronoms personnels sont aussi fortement sujets au phénomène de contraction (fusion dextroverse, le pronom s'attachant à l'élément précédent : *si+me* → *sim*, *jo+le* → *jol*, etc.), l'élision des pronoms personnels est souvent traitée non comme une élision mais comme le pendant sinistroverse de la contraction. Ainsi, dans la GGHF, l'élision des pronoms n'est pas une élision mais une proclise ; LUQUIENS (1919 [1909], p. 83-84, 88) traite de (*il*) *me+aime* → (*il*) *m'aime* comme d'une contraction, et (WALKER 1981, p. 51) traite les élisions préverbaux comme des syncopes, mais parle bien d'élisions pour les cas devant préposition (*Eulalie* : *ell'ent*). Ce choix théorique, parfois inconscient, se retrouve dans la façon dont les éditeurs rendent les segments en question : les contractions n'ont en principe pas d'apostrophe (cf. *lint* en 9.15), tandis que les élisions en ont une (*Eulalie* : *ell'ent*). Il importe donc, ici plus que jamais, de ne donner aucune valeur à la présence ou à l'absence d'une apostrophe éditoriale dans l'analyse d'une forme.

LUQUIENS (1919 [1909]). *An introduction to old french phonology and morphology.*

WALKER (1981). *An Introduction to Old French Morphophonology.*

En somme. Alors que les pronoms régimes connaissent d'emblée les deux paradigmes, forts et faibles, et que leur paradigme faible s'élide dès les premiers textes et que cette élision revêt un caractère obligatoire, ce n'est pas le cas des pronoms sujets de 1-2sg, dont les séries fortes et faibles ne semblent se séparer qu'au XII^e s. et dont la série faible connaît une élision qui n'est d'abord que facultative.

9.1.2 Traitement des données

Nous ne nous sommes intéressé qu'aux pronoms personnels se terminant par une voyelle, c'est-à-dire les pronoms de première personne singulier aux cas sujet et régime (*je*, *jou*, *me*, etc.), de deuxième personne singulier aux deux cas également (*tu*, *te*), et de troisième personne du singulier, uniquement au cas régime au masculin (*le*), et aux deux cas au féminin (*ele*, *la*). Cette recherche serait évidemment consolidée par l'inclusion des pronoms non personnels, mais un certain nombre d'entre eux sont annotés en partie du discours dans le corpus d'une manière qui rend peu aisé leur traitement par Comparalem (cf. *que* en sec. 10.1.2), tandis que d'autres disposent de fréquences beaucoup moins élevées. Les pronoms personnels représentent la catégorie à annotation en pos fiable la mieux attestée en termes de fréquence.

Étant données les difficultés patentes de Tree Tagger pour lemmatiser ces formes,¹⁵ nous nous sommes fondé sur l'annotation des parties du discours de Dees et avons créé des étiquettes de lemme *ad hoc*, reprenant simplement les caractéristiques de personne et de nombre. Nos lemmes sont donc PRO_1sg, PRO_2sg et PRO_3sg. Enfin, nous avons écarté manuellement 297 occurrences, soit parce qu'elles se terminent par un <a> final atone au lieu d'un <e> (25 occ. *ela* pour *ele*), soit parce qu'en contexte elles ne sauraient correspondre à un pronom personnel.

Pour les pronoms régimes de 2-3sg, nous avons pu isoler les formes atones des formes toniques, les deux séries n'étant pas homographes (cf. table 9.1); nos données ne portent alors que sur les séries atones. Nous avons également exclu les formes au cas régime indirect. En revanche, pour les pronoms sujet et régime de 1sg ainsi que pour le pronom sujet de 2sg, une telle dissociation n'a pas été possible. Pour ces pronoms, nous avons exclu les formes qui sont spécifiquement toniques (*gié, mei, tei*, etc.), mais nos données comportent de nombreuses formes qui appartiennent aux deux séries. La lecture proposée *infra* entre ces deux types de données est donc différente.

Dans les contextes interrogés, la quantité de forme retenues va de 4739 occurrences pour le PRO_PERS_FEM_CR_3sg (*la, le, l*) à 27 385 occurrences pour le PRO_PERS_MASC_CR_3sg (*le, l*). Les deux autres catégories les mieux attestées sont celles pour lesquelles nous n'avons pas pu appliquer de restriction totale contre les formes toniques, à savoir le PRO_PERS_1CSsg (18 986 occ. *je, ge, jo, jeo, jou, j, g*) et le PRO_PERS_2SGCS (16 098 occ. *tu, te, t*).

9.2 Présentation des résultats

Puisque nous ne travaillons pas, ici, sur une large catégorie, mais sur un ensemble défini et restreint de formes au sein d'une micro-catégorie, la présentation des résultats suit un ordre différent. La table 9.2 présente un résultat des taux de défaut de voyelle finale atone pour les pronoms retenus, selon le cas et le genre. Elle précise, par ordre de colonne, le nombre total d'occurrences (dans les contextes retenus), puis le nombre d'occurrences (avec et sans voyelle finale atone) devant initiale consonantique et le taux de formes défectueuses parmi ces formes, avant d'indiquer les mêmes données pour les occurrences prévocaliques.

TABLE 9.2 – Quantité de formes et taux de formes en défaut de <e> pour les pronoms personnels.

Personne	Cas	Nb.	_#C		_#V	
			Nb.	Tx_Ø	Nb.	Tx_Ø
1sg	CS	18 986	15 794	0,01%	3192	45,9%
	CR	16 098	9746	0,14%	7352	97,61%
2sg	CS	6430	4944	<0,01%	1486	1,21%
	CR	5054	3207	0,16%	1847	96,16%
3sg MASC	CR	27 385	13 757	18,5%	13 628	98,58%
3sg FÉM.	CS	8250	5888	12,72%	2362	6,44%
	CR	4739	4340	0,05%	339	33,33%

15: À titre d'exemple, l'ensemble des lemmes de Tree Tagger correspondant en réalité à des occurrences du pronom personnel de 1sg, et que nous avons fusionnés sous l'étiquette PRO_1sg, est : *je, je-moi, jou, jeu-juene, mon1, mon1-mon, jesir, mai-mui, mi2-mi, mien, jöa12-jöe, mai*.

Comme mentionné *supra*, nous n'avons pas pu trier entre formes atone et tonique pour le pronom sujet de 2sg, et n'avons pu trier que partiellement pour le pronom sujet de 1sg, tandis que la distinction entre formes toniques et formes atones ne s'applique pas au pronom sujet de 3sg FÉM. Il est donc naturel que ces catégories disposent des taux de défaut les moins élevés, avec, devant initiale vocalique, 1% de défaut pour le pronom de 2sg CS, 6% de défaut pour le pronom de 3sg FÉMCS. Il en va de même pour le taux moyennement élevé du pronom sujet de 1sg, avec 46% de défaut. Dans tous ces cas, les centiles qui séparent la valeur du taux de la valeur de 100% correspondent à un mélange de formes toniques (qui, donc, ne sont pas sujettes à l'élision), et de formes atones (qui, donc, devraient être sujettes à l'élision).

9.2.1 Trouble dans le genre des PROPERsCR3sg

Plus étonnant est le taux du pronom de 3sg FÉMCR (*la, le* (pic.), *l*), dont seulement 33% des formes antévocaliques sont en défaut, alors que nous avons pu, ici, isoler les formes atones, et que les formes toniques ne viennent donc pas parasiter les résultats. L'une des raisons de la faiblesse de ce résultat est que de très nombreux pronoms élidés, réduits à la forme *l*, ont été annotés comme masculins dans le corpus, alors que leur référent est féminin. Ainsi, en (9.17), le pronom élidé désigne *la vache*, mais est annoté comme masculin. De même, en (9.18), le pronom élidé renvoie à *la reine* (mentionnée à la phrase suivante), mais est annoté comme masculin.

(9.17) *vient la vache seule li chiens salt contre la vache si l assalt et chace*¹⁶
« la vache vient seule, le chien lui saute dessus, l'assaille et la chasse »
(ailed, p. 9, trad. pers.)

16: *l* est annoté pos = "PRO_pers_obj_-masc_3_sg.

(9.18) *en une cambre l a menee*¹⁷
« en une chambre il l'a emmenée » [la reine]
(bodo, p. 16, trad. *id.*)

17: *l* est annoté pos = "PRO_pers_obj_-masc_3_sg.

De fait, à la troisième personne, le taux de pronoms régimes masculins est surestimé, tandis que celui des pronoms régimes féminins est sous-évalué. Néanmoins, cette surestimation ne rend pas tout à fait irrecevable le taux des pronoms masculins. On peut en effet postuler que le rapport entre occurrences préconsonantiques et prévocaliques doit être équivalent entre les deux genres. Si l'on accepte ce postulat, le nombre d'occurrences manquantes au féminin est d'environ 2000. Pour les PROPERsMASC3sg, sur les 13 628 occurrences, 13 435 sont élidées (*l*) et 193 seulement ne le sont pas; en retirant 2000 occurrences qu'on suppose féminines, seules les décimales du taux évoluent (98,34% au lieu de 98,58%). Et, même si notre estimation d'environ 2000 occurrences féminines élidées annotées comme étant masculines était insuffisante, il s'en faudrait de beaucoup pour que le taux du masculin change drastiquement. On peut donc considérer, malgré la présence de formes féminines dans les données masculines, que le taux des PROPERsMASC3sg est plutôt fiable.

Quant aux PROPERsFÉMCR3sg, le rétablissement de 2000 occurrences élidées porterait le taux de défaut devant initiale vocalique à 88,91%, avec 2133 formes élidées pour 266 formes *la* et *le* devant initiale vocalique. Ce taux, s'il n'égale pas celui des PROPERsMASC3sg, s'en rapproche.

L'estimation de 2000 formes mal annotées en genre, qui découle du simple postulat que le rapport entre nombre d'occurrences préconsonantiques et prévoaliques n'a pas de raison *a priori* d'être différent au masculin et au féminin, est donc confortée par le fait qu'elle rapproche le taux d'élision des formes féminines *la, le* (pic.) de celui des formes masculines (*le*). Dans les deux cas, la marge d'erreur d'une telle manipulation à l'aveugle est considérable; il ne s'agit donc pas de corriger précisément les taux, mais simplement de considérer que le taux réel des PRO_{PERS}FÉM_{CR}3sg se situe plus probablement autour de 90% de défaut de <e> qu'autour de 33%. Puisque le taux mesuré de 33% est lourdement sous-estimé, mais en l'absence d'une méthode fiable pour corriger ce taux et obtenir une mesure définitive, nous reproduisons les données putatives des PRO_{PERS}FÉM_{CR}3sg entre parenthèses et précédées d'un signe d'approximation '~', tandis que les données des PRO_{PERS}MASC_{3sg} seront reproduites précédées du même signe mais sans parenthèses, étant donné qu'elles semblent peu sensibles au retrait de quelques milliers d'occurrences défectueuses.

Si une erreur d'annotation analogue peut sembler être suggérée par la différence radicale de taux entre le CS et le CR du PRO_{PERS}2sg (1,21% vs 96,16%), nous n'avons pas relevé de telle erreur systématique dans les données, et le rapport entre nombre d'occurrences préconsonantiques et prévoaliques ne plaide guère en faveur d'une telle erreur. De même, il ne semble pas y avoir lieu de corriger le taux des PRO_{PERS}CS_{3sg} (6%).

Se dessine alors une asymétrie forte entre les cas sujets et les cas régimes, les premiers étant bien moins souvent élidés que les seconds. Si cette asymétrie tient au moins en partie au fait que nous n'avons pas pu isoler les formes atones pour les PRO_{PERS}CS_{1-2sg} alors que nous avons pu le faire pour leur cas régime, elle ne tient peut-être pas uniquement à ce facteur, ce qui induirait une situation analogue à celle du français contemporain, avec une élision informelle et non systématiquement transcrite au cas sujet (*t'aimes ?*), et une élision non connotée en termes sociolinguistiques et toujours transcrite au cas régime (*il t'aime*).

9.2.2 Typologie

Les données en table 9.2 *supra* évoquent une typologie à trois entrées. La première de ces entrées tient, comme on l'a dit, au traitement des données, selon que les formes atones sont homographes de tout ou partie des formes toniques, et/ou qu'il existe une série non spécifiée quant à son statut accentuel. Ainsi, les lemmes pour lesquels une exclusion de toutes les formes toniques n'était pas possible se caractérisent par des taux de défaut devant initiale vocalique inférieurs à ceux des catégories dont les formes toniques sont systématiquement exclues (9.19).

(9.19) Typologie par isolement des formes atones (__#V)

(a) Les formes atones sont isolées

.i. 1sg CR	<i>me</i>	98% de <i>m</i>
.ii. 2sg CR	<i>te</i>	96% de <i>t</i>
.iii. 3sg MASC CR	<i>le</i>	~99% de <i>l</i>
.iv. (3sg FÉM CR	<i>la, le</i>	~90% de <i>l</i>)

(b) Les formes atones ne sont pas isolées

.i. 1sg CS	<i>je, jo, jeo, jou, je, ge</i>	46% de <i>j, g</i>
.ii. 2sg CS	<i>tu, te</i>	1 % de <i>t</i>
.iii. 3sg CS	<i>ele</i>	6% de <i>el</i>

Le deuxième critère typologique tient au rapport entre les taux de défaut devant initiale vocalique et ceux devant initiale consonantique. En effet, pour tous les pronoms personnels sauf le pronom de 3sg CS, le taux de défaut devant initiale vocalique est substantiellement plus élevé que le taux devant initiale consonantique, avec une différence comprise entre 97% (9.20 a.i.) à 1% (9.20 b.ii.). À l'inverse, pour le PRO_{PERS}FÉMCS3sg, le taux de défaut est 6% plus élevé devant initiale consonantique que devant initiale vocalique (9.20 c). La typologie, ici, peut être affinée, comme illustré en (9.20), en séparant le PRO_{PERS}CS2sg des autres pronoms dont le taux de défaut devant initiale vocalique est supérieur au taux devant initiale consonantique, dans la mesure où le taux de défaut devant initiale vocalique de *tu, te* sujet est très faible dans les deux cas, ce qui témoigne d'une résistance globale à la perte du contenu segmental. La position du pronom de 1sg CS au sein du continuum est alors en partie arbitraire, dans la mesure où la différence entre ses taux contextuels se situe à mi-chemin entre les deux groupes; en (9.20), ce pronom est classé avec celui de 2sg CS, pour des raisons expliquées *infra*.

(9.20) Classement par différence contextuelle des taux de défaut

	__#V	__#C	Différence
(a) \emptyset __#V \gg \emptyset __#C			
.i. 1sg CR	98%	0,14%	97%
.ii. 2sg CR	96%	0,16%	96%
.iii. (3sg FÉM CR	~90%	0,05%	90%)
.iv. 3sg MASC CR	~99%	19%	80%
(b) \emptyset __#V > \emptyset __#C			
.i. 1sg CS	46%	0,01%	46%
.ii. 2sg CS	1%	<0,01%	1%
(c) \emptyset __#V < \emptyset __#C			
.i. 3sg FÉM CS	6%	13%	-6%

Le classement du pronom de 1sg CS en (9.20 b) et non en (9.20 a) peut paraître arbitraire du point de vue de la simple différence entre les taux de défaut telle que reportée en (9.20); son taux ambigu provient en partie de la diachronie, dans la mesure où ce pronom ne connaît pas de défaut devant initiale vocalique au début du corpus. Néanmoins, il faut remarquer que le nombre d'occurrences défectueuses devant initiale consonantique du pronom de 2sg CS est si faible que le rapport de division entre son taux d'occurrences défectueuses devant initiale vocalique et son taux d'occurrences défectueuses devant initiale consonantique est

supérieur à 100. Par ailleurs, ces deux pronoms forment, avec le pronom féminin de 3sg CS en (9.20 c), les trois catégories pour lesquelles nous n'avons pas été en mesure d'isoler les formes atones (premier critère : 9.19 b). Il y a donc une interaction directe entre nos deux premiers critères typologiques : les pronoms pour lesquels la séparation entre formes atones et toniques n'était pas possible sont ceux dont le rapport (de soustraction, non de division) entre les taux contextuels est le moins élevé, et également ceux dont le taux de défaut devant initiale vocalique est inférieur à 50%. Néanmoins, cette interaction ne saurait suffire à expliquer ces taux, dans la mesure où il semble nécessaire de couper le continuum des taux entre d'une part le pronom de 1sg CS (46% $_ \#V$) et d'autre part les pronoms de 2sg CS (1%) et de 3sg FÉM (6% $_ \#V$), et où l'inversion du rapport entre les taux contextuels ne saurait s'expliquer par le critère d'identification des formes atones.

Le troisième critère typologique est phonologique, séparant les pronoms selon que la voyelle susceptible de s'élider est un schwa ou une autre voyelle, c'est-à-dire selon que toutes les formes atones du pronom sont en <e> ou non. Ainsi, seul un schwa est susceptible de disparaître pour les pronoms de 1sg CS-CR (*je, ge; me*),¹⁸ 2sg CR (*te*) et de 3sg CR masculins (*le*), CS féminin (*ele*), ainsi que dans la forme picarde du pronom du CR féminin *le* (9.21 a). En revanche, la forme sujette à la perte de sa voyelle finale comporte soit un <e> soit une autre voyelle pour le pronom de 2sg CS (*tu, te*) et les formes archaïques du CS de la première personne (*jo, jeo, jou* : 9.21 b), si l'on admet avec MARCHELLO-NIZIA (2015) que les cas d'élisions requises par la métrique mais non transcrites dans la forme graphique du mot attestés dans *Roland* témoignent de la possibilité de l'élision du pronom de 1sg CS (9.21 b.ii.); dans le cas de *Roland*, on compte 132 formes *jo* pour 8 formes *je* de ce pronom.¹⁹ Enfin, pour le pronom CR féminin, ce n'est qu'en picard que la forme susceptible de se voir amputer de sa voyelle peut être *le*; dans les autres dialectes, seul *la* et *l* sont susceptibles de représenter ce pronom, ce qui signifie que la voyelle manquante ne saurait être un schwa (9.21 a.vi.).

(9.21) Classement par voyelle sujette à effacement

- (a) /ə/ → Ø
- | | |
|------------------------|--------------------------------|
| .i. 1sg CS | <i>je, ge</i> |
| .ii. 1sg CR | <i>me</i> |
| .iii. 2sg CR | <i>te</i> |
| .iv. 3sg MASC CR | <i>le</i> |
| .v. 3sg FÉM CS | <i>ele</i> |
| .vi. 3sg FÉM CR (pic.) | <i>le</i> (pic.) ²⁰ |
- (b) {/ə/, V} → Ø
- | | |
|-----------------------|-------------------------|
| .i. 2sg CS | <i>tu, te</i> |
| .ii. 1sg CS arch. (?) | <i>je, jo, jeo, jou</i> |
- (c) V → Ø
- | | |
|---------------------------|-----------|
| .i. 3sg FÉM CR (non pic.) | <i>la</i> |
|---------------------------|-----------|

Cette troisième catégorisation ne recoupe que partiellement les deux autres catégorisations présentées ci-dessus. En effet, si l'on s'en tient aux résultats tels que Comparalem les propose, ce sont les pronoms dont les taux de défaut sont les plus élevés qui sont ceux dont la voyelle sujette à effacement est toujours un schwa (1sg CR *me* > 2sg CR *te* > MASC CR 3sg

18: Sauf, en 1sg CS, pour les formes archaïques, cf. 9.21 b.ii. *infra* et sec. 9.1.1.1 *supra*. Nous faisons ici l'hypothèse que dans *je, ge* atones, <e> représente un schwa, ce qui ne saurait être le cas dans *je, ge* toniques.

19: Et, comme illustré en (9.3) - (9.11) *supra*, ce ne sont pas les formes *je* qui sont suspectées d'élision métrique.

20: Il n'y a pas de raison de classer le dialecte picard sous (9.21 b), étant donné que les formes picardes *le* sont pratiquement exclusives des formes 'non picardes' *la* au sein des textes picards. Selon GOSSEN (1970, p. 122), la forme *la* pour *le* est « extrêmement rare » dans les textes picards, qu'ils soient littéraires ou documentaires. Ce trait est confirmé par les données de DEES (1980, c. 26, p. 28), qui obtient des taux de 100% de formes *le* pour le PRO-PERS-FÉM CR 3sg dans le Hainaut, le Nord et la région Somme, Pas-de-Calais — la carte de Dees est reproduite en fig. E.6 en annexes, p. 615. En revanche, aux marges de la picardie linguistique, la carte de Dees justifierait un classement sous (9.21 b), avec 86% de formes *la* dans l'Aisne, et 63% dans l'Oise.

le ≤ 80% __#V). Néanmoins, cela ne concerne pas le pronom sujet de FÉM3sg *ele* (13% __#V), et fait abstraction du problème d'annotation en genre des pronoms de FÉMCR3sg *la*, dont nous estimons que le taux réel est proche de 90%. La présence d'un schwa final n'est donc ni une condition suffisante (*ele*) ni, si l'on accepte la correction des données pour *la*, une condition nécessaire à l'obtention de taux de défaut élevés devant initiale vocalique.

Les données qui seront détaillées *infra* montrent que l'identité (schwa ou autre voyelle) de la voyelle finale atone n'a que peu d'importance. Certes, ce sont des formes en schwa final qui montrent les plus forts taux de défaut, mais c'est parce que ce sont les formes de CR qui sont en schwa, tandis que les formes de CS soit ne sont pas systématiquement en schwa, soit présentent des confusions ou homographies entre série atone et série tonique; le critère prédominant est donc celui du cas, et non de l'identité de la voyelle. Quant au pronom féminin de 3sg CR *la*, les données nous manquent puisque, comme précisé dans la section précédente, les occurrences *l* féminines sont pour la plupart annotées comme étant masculines; il y a fort à parier que le taux de défaut de *la* n'est pas différent de celui du masculin *le*. Par ailleurs, des cas sujets dotés de formes en <e> comme le pronom sujet de 2sg *tu, te* ne connaissent que très faiblement le défaut de voyelle finale atone, ce qui est un élément de plus pour discréditer le rôle de l'identité segmentale de la voyelle. Il est donc naturel que la typologie phonologique ne recoupe que partiellement les deux autres typologies proposées.

9.2.2.1 Défauts devant #C

Comme indiqué en table 9.2 *supra*, tous les pronoms disposent également d'occurrences défectueuses devant initiale consonantique et, comme vu en (9.20) avec le deuxième critère typologique proposé, le taux d'occurrences préconsonantiques défectueuses est bien plus faible que le taux d'occurrences préconsonantiques défectueuses, et est même marginal dans bien des cas. Les seules exceptions à ce constat concernent les pronoms de 3sg MASC CR *le* et FÉMCS *ele*. Pour *le*, le taux de défaut devant initiale consonantique (19%) reste bien inférieur au taux devant initiale vocalique (~99%) et est le plus élevé des taux préconsonantiques. Pour *ele*, le taux de défaut devant initiale consonantique reste bien plus élevé que celui des autres pronoms (13% > 0,16% pour 2sg CR *te* > tous les autres pronoms) et, surtout, il est plus élevé que le taux de défaut devant initiale vocalique (6%).

Enclise. Pour l'ensemble des pronoms, ces défauts devant initiale consonantique représentent des cas d'enclise : le pronom personnel atone, placé immédiatement après un autre élément atone, fusionne avec celui-ci. Les éléments support d'enclise repérés dans la bibliographie²¹ sont les prépositions *a* (9.22), *de* (9.23), *en* (9.24), l'élément *si/se* (9.25), l'adverbe ou la conjonction *ne* (9.23), le pronom personnel 1sg CS (9.26), et le pronom conjonctif *qui/que* (9.27). Dans la plupart des éditions et grammaires, ces enclises sont transcrites d'un seul mot (*al, nel, jel, nel, sel, quel, quil*), mais ça n'est le cas que pour une minorité de formes dans le NCA, et certains éditeurs préfèrent l'usage de l'apostrophe, comme SHORT (1990) qui édite les occurrences en deux mots, le second étant suivi

21: Peu de grammaires donnent une vision d'ensemble des cas d'enclise; le phénomène est la plupart du temps mentionné au titre des changements phonétiques qu'il implique. C'est le cas p.ex. chez FOUCHÉ (1969 [1958], p. 529, 440, 1966, p. 653, 859, 866), CLÉDAT (1917, p. 64) et PRICE (1985 [1971], p. 115-116); c'est le cas pour certaines enclises chez É. BOURCIEZ et J. BOURCIEZ (1974 (1967), p. 45-46). En revanche, LA CHAUSSÉE (1989 [1974], p. 120-121), NYROP (1899, p. 239-240) et POPE (1966 [1934], p. 322-323) y consacrent une section, et c'est chez ces auteurs qu'on trouve les listes les plus complètes, bien qu'elles divergent à la marge.

d'une apostrophe même devant initiale consonantique (*que l' blasme et qui l' lot* en 9.27). Néanmoins, le NCA édite parfois les formes en un seul mot, comme *enl* en (9.24) ; cette pratique, plus répandue chez les éditeurs, permet une édition homogène des formes où l'enclise est transparente et des formes dont le pluriel et/ou l'évolution phonétique rend opaque l'enclise originelle (p.ex. *a+les* → *as, aus* ; *de+le* → *dou, du* ; *si+le* → *sis* ; etc.).

- (9.22) *je n'ai droit a l retenir*
« je n'ai pas le droit de le [l'anneau] garder »
(*ombre*, p. 17, trad. pers.)
- (9.23) *il dist qu il ne l cuidera james, car ceste aventure l en a mis tot fors de l cuidier*
« il dit qu'il ne le croira jamais, car cette aventure l'a disposé à tout sauf à le croire » (*queste*, p. 18, trad. *id.*)
- (9.24) *beneistrai tei en ma vie enl tuen num leverai mes palmaes*
« je te bénierai en ma vie, en ton num [= en le tien nom] je lèverai mes mains » (*psautier*, p. 50, trad. *id.*)
- (9.25) *au prouost vint se l salua et il fierement l apiela*
« il alla vers le prévôt, le salua et lui adressa fierement ces paroles » (*fabj*, p. 6, trad. *id.*)
- (9.26) *dist li hermites 'je l te donrai*
« L'hermite dit : « je te le donnerai » (*aiol*, p. 12, trad. *id.*)
- (9.27) *ambure ocit ki que l blasme ne qui l lot*²²
« [il] les tue tous deux, qu'on le blâme ou non » (*Roland*, v. 1589, p. 40, trad. *SHORT* 1990)

Tous les exemples en (9.22 - 9.27) sont donnés avec le *ProPersCR3sg* ; c'est le seul qui permet l'enclise avec tous ces supports. Les enclises avec d'autres seconds éléments que le *ProPersCR3sg* sont essentiellement données comme archaïques. Elles concernent le *ProPersCR1sg* (*si+me* → *sim* dans *Roland*,²³ *ne+me* → *nem*²⁴), le pronom réfléchi singulier (*si+se* → *sis* dans *Jonas*).²⁵ Dans cette période archaïque, l'enclise est également possible après un mot fonctionnel autre que ceux listés *supra*, avec *poro+se* → *poros* (*Eulalie*), *sempre+lo* → *semprel* (*Jonas*), *lui+les* → *luis* (*Vie de saint Léger*), *porquei+me* → *porqueim* (*Alexis*) (*NYROP* 1899, p. 239-240), *contre+le* → *contrél* (uniquement __#C, *POPE* 1966 [1934], p. 323), *ja+te* → *jat* (*WALKER* 1981, p. 51). *POPE* (1966 [1934], p. 323) et *WALKER* (1981, p. 51) notent également quelques cas avec des mots lexicaux en support : *destre+le* → *destrel*, *faire+le* → *faïrel* (*Pope*) ; *terre+le* → *terrel*, *chambre+le* → *chambrel* (*Walker*), et *Pope* précise que ces exemples sont tardifs et limités au Sud-Ouest.

Cette restriction générale de l'enclise aux seuls pronoms de 3sg explique leurs taux de défaut préconsonantiques bien supérieurs à ceux des autres pronoms (Masc 19%, Fém 13%), sachant que de nombreuses formes féminines *l* sont annotées comme masculines. Les autres formes défectueuses devant #C sont en quantité si marginales qu'il est difficile de savoir quel statut leur accorder. Certaines occurrences tiennent visiblement à des erreurs dans la source ou dans l'édition. Ainsi, trois occurrences des six occurrences *j* devant initiale consonantique proviennent du *Perceval* de Chrétien de Troyes, dont le NCA dispose de deux versions (*perb* et *perr*), dont l'une est dupliquée (*percevalb* correspond à *perb*). Ces occurrences sont placées au début d'un vers, ce qui interdit en principe

ombre : JEAN RENART (1221*). *Le Lai de l'Ombre*. Dees : [Marne] ; compo. pic., ca 1221 ; ms. frc, ca 1288 ; CR1 (JEAN RENART 1913) ; DEAF : *OmbreB*¹.

queste : ANON. (1228*). *La queste del saint Graal*. Ms. K, Dees : Nièvre, Allier ; compo. nil, ca 1228 ; ms. nil, ca 1250 ; CR3 (PAUPHILET 1923) ; DEAF : *SGraalivQuestP*.

psautier : ANON. (1125*). *Livre des psaumes*. Dees nil ; compo. a.n., ca 1125 ; ms. a.n., ca 1150 ; CR2 (MICHEL 1876), DEAF : *PsCambrM*.

fabj : ANON. (1250*). *Fabliaux n°2*, 4. Ms. J ; Dees : Somme, Pas-de-Calais ; compo. pic., ca 1250 ; ms. pic., ca 1285 ; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).

aiol : ANON. (1160*). *Aiol*. Dees : Somme, Pas-de-Calais ; compo. pic., ca 1160 ; ms. pic., ca 1275 ; MS1 (NORMAND et RAYNAUD 1877) ; DEAF : *Aiol*^{1/2N}.

22 : La BFM (MOIGNET 1972a) édite *quel et quil* ; *SHORT* (1990) édite *que l' blasme et qui l' lot*.

rolandox : ANON. (1100*). *La Chanson de Roland*. Ms. O ; Dees : Angleterre ; compo. N.-O., ca 1100 ; ms. a.n., ca 1137 ; CR1 (SEGRE 1971) ; DEAF : *RoLS*.

23 : *NYROP* 1899, p. 240 ; *WALKER* 1981, p. 51.

24 : *NYROP* 1899, p. 240.

25 : *NYROP* 1899, p. 240 ; É. BOURCIEZ et J. BOURCIEZ 1974 (1967), p. 45-46 ; *WALKER* 1981, p. 51.

toute enclise avec le terme précédent. Les ms. présents dans le NCA, siglés B et R/P8 dans la tradition, semblent bien porter la graphie *i*; le ms. P, édité par KUNSTMANN (CHRÉTIEN DE TROYES 2009a) et repris par la BFM, porte *qui* à la place de *i*, et POIRION (CHRÉTIEN DE TROYES 1994), dont le ms. de base est aussi P, corrige ce *qui* en restituant *i* à partir de R/P8 (le ms. de perr). Le <j> du NCA est donc un <i>, c'est-à-dire l'adverbe de lieu ou, éventuellement, la forme sans <l> du pronom personnel sujet de 3sg masculin (<i> pour <il>). Il ne saurait, en aucun cas, être un pronom personnel sujet de première personne, et ce sont l'édition et l'annotation du NCA qui sont fautives.

- (9.28) NCA *li rois artus que ueoir uoil
j fait cheualiers ce dit len*
BFM *Le roi Artus, que veoir uuel,
qui fet chevaliers, ce dit an.*
POIRION *Li²⁶ rois Artus, que veoir uuel,
I²⁷ fet chevaliers, ce dit an.*
« Le roi Arthur, que je veux voir, y fait des chevaliers, à ce qu'on dit. »
(Perceval; NCA : perb & percevalb p. 17, perr p. 16; BFM PercevalKu v. 838; POIRION : CHRÉTIEN DE TROYES 1994 v. 840; trad. POIRION *id.*)

Ele ~ *el*. Si l'enclise est en mesure d'expliquer une part conséquente des défauts de <e> devant initiale consonantique du pronom de 3sg CR *le*, il n'en va pas de même pour le pronom de 3sg CS *ele*. La distribution entre formes bisyllabiques (type *ele* : *ele, elle, ela, ille*)²⁸ et monosyllabiques (type *el* : *el, il, ell*) montre une claire prédominance du type bisyllabique, légèrement moins marquée devant initiale vocalique (6% de *el*) que devant initiale consonantique (13% de *ele*). Par ailleurs, aucune grammaire ne fournit d'exemple d'enclise avec ce pronom sujet, et contrairement aux autres pronoms sujets à défaut de <e>, il est le seul à être bisyllabique.

Les formes du PROPERSFÉMCS3sg du type *el* identifiées par Comparalem et dont la fréquence est supérieure à 5 occ. sont *el* (1055 occ.), *il* (23 occ.), *ilh* (7 occ.)²⁹ et *ell* (6 occ.). Comme il sera abordé *infra*, ces formes ne montrent pas une répartition diatopique très tranchée, même si elles sont un peu plus nombreuses dans la moitié Ouest du domaine. Il ne s'agit donc pas non plus d'une variation diatopique. Les formes *el* semblent donc précéder d'une autre logique, et nous nous en tenons à admettre que la variation entre le type *ele* et le type *el* est d'un autre ressort que les variations étudiées ici.

9.2.2.2 En somme

Les pronoms personnels sujets sont donc susceptibles de connaître des défauts de leur voyelle finale atone. Ces défauts sont largement majoritaires devant initiale vocalique (< 90%) pour les pronoms dont nous avons pu isoler les formes atones, sont légèrement minoritaires (1sg CS : 46%) ou très minoritaires (≤ 6%) pour les autres formes. Par ailleurs, pour tous les pronoms sauf celui de 2sg CS et celui de 3sg FÉMCS, la quantité de défauts est bien plus élevée devant initiale vocalique que devant initiale consonantique. De manière orthogonale, la voyelle

26: « Le P. Nous corrigeons d'après P12. »

27: « Qui P. Nous corrigeons avec Hilka d'après P12. »

perb : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS B; Dees : nil; compo. champ. mérid., ca 1180; ms. bourg. sept., ca 1310; MS1 (transcr. inédite R. L. H. Lops); Dupliqué : percevalb.

percevalb : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS B; Dees : Haute-Marne; compo. champ. mérid., ca 1180; ms. bourg. sept., ca 1310; MS1 (transcr. inédite R. L. H. Lops); Dupliqué : perb.

perr : CHRÉTIEN DE TROYES (1180*). *Perceval*. Dees : [Hainaut]; compo. champ. mérid., ca 1180; ms. champ., ca 1237; MS1 (transcr. inédite R. L. H. Lops); inclus partiellement dans NCA_chret1.

PercevalKu : CHRÉTIEN DE TROYES (1183*). *Conte du Graal (Perceval)*. Compo. ca 1181-1185, ms. ca 1235, dial. champ. (CHRÉTIEN DE TROYES 2009a).

28: Si, dans l'*Atlas des chartes*, DEES (1980, p. 26, c. 24) trouve des formes en <i> initial dans plusieurs régions du Nord-Est (région Moselle, Meurthe-et-Moselle, Vosges, Aisne, Hainaut, Wallonie et Ardenes), ces formes se limitent ici à *ille* dans quatre textes : bern2 (32 occ.), malk, miro et perf (1 occ. chacun). La carte de Dees est reproduite en annexes, en fig. E.7 (p. 616).

bern2 : BERNARD DE CLAIRVAUX (1190*). *Sermans*. Dees : Moselle, Meurthe-et-Moselle; compo. lorr., ca 1190; ms. lorr., ca 1190; MS (BERNARD DE CLAIRVAUX 1894); DEAF : SBernAn2S.

malk : JEHAN MALKARAUME (1283*). *La bible de Jehan Malkaraume*. Dees : Marne; compo. lorr., ca 1283; ms. lorr., id.; CR2 (JEHAN MALKARAUME 1987); DEAF : BibleMalkS.

miro : ROBERT DE L'OMME (1266). *Le miroir de vie et de mort*. Ms. G; Dees : Somme, Pas-de-Calais; compo. pic., 1266; ms. art., 1277; MS1 (LÅNGFORS 1921, 1924); DEAF : RobOmMirL.

perf : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS F; Dees : nil; compo. champ. mérid., ca 1180; ms. Est, ca 1250; MS1 (transcr. inédite R. L. H. Lops).

29: *ilh* est la forme wallonne qui correspond en principe au pronom masculin sujet; l'existence de formes féminines en <i> initial dans la région (cf. note E.7) permet de considérer que les occurrences *ilh* annotées comme féminines représentent la variante 'doublement wallonne' (<i> pour <e> et insertion de <h>) du type *el*.

susceptible d'être déficiente est le plus souvent un schwa, mais ce critère n'est pas une condition suffisante et ne semble pas non plus être une condition nécessaire. Les défauts devant initiale consonnantique viennent principalement du phénomène de l'enclise (pour *le* → *-l*) et, pour *el*, d'une variation qui semble plus libre; pour les autres formes, ils sont véritablement marginaux et tiennent pour la plupart à des problèmes d'édition et/ou d'annotation du corpus.

Dans ce qui suit, nous étudions l'impact de différentes variables pour chaque pronom, par ordre décroissant de taux de défaut devant initiale vocalique.

9.2.3 PROPersMAscCR3sg : *le* ~ *l*

Le pronom personnel régime masculin de troisième personne du singulier, dont la forme atone en construction directe est *le*, est sujet à défaut de son <e>, par élision (*l* + mot suivant) ou par enclise (mot précédent + *l*). Dans ce qui suit, nous traitons des cas de défaut de <e> de ce pronom, d'abord en fonction du contexte et de la diachronie, puis en fonction de la diatopie et de l'intersection entre diatopie et diachronie.

9.2.3.1 Distribution contextuelle et diachronique

Les graphiques de séries temporelles en fig. 9.1.a-d, en page suivante, représentent le taux de défaut de <e> de chaque texte pour les pronoms personnels régimes de troisième personne du singulier (type *le* ~ *l*). Le graphique en (e) sur la même figure illustre la différence entre les deux contextes, mesurée par le test de Student en (f). La différence entre le contexte prévoicalique et le contexte préconsonnantique est frappante, mais ne saurait être caractérisée comme une absence de défauts devant initiale préconsonnantique.

***l* __#C.** Le taux moyen de *l* devant initiale consonnantique est de 21%, et certains textes disposent d'un taux de 100% de défaut de <e> dans ce contexte. Les 2549 occurrences *l* (pour le MAscCR3sg) devant initiale consonnantique correspondent bien, à une écrasante majorité, à des occurrences d'enclise répertoriées dans les grammaires. Toutes celles qui sont dotées d'une fréquence supérieure à 40 sont données en table marginale 9.3. Ainsi, l'enclise dont le support est *ne* représente plus de la moitié des occurrences préconsonnantiques où le pronom se réduit à *l*, et les trois enclises les plus courantes (*nel*, *sel* et *jel*) représentent 92% de ces cas. Sous 'autres' en table 9.3, on retrouve des configurations dont le terme support d'enclise est un monosyllabe fonctionnel (52 occ., p.ex. *ce l*, *de l*, *tu l*) et quelques cas dont le terme support d'enclise est un mot lexical, monosyllabique ou polysyllabique (22 occ., p.ex. *ainzque l*, *feire l*, *ja l*). Si certaines occurrences peuvent ponctuellement se révéler suspectes — par exemple lorsque le terme support se termine par une consonne; *comment l*, *quequ l*, *prophetes l* — ces cas ne concernent qu'une infime minorité des enclises relevées.

Diachroniquement, la tendance à la décroissance de l'enclise évoquée par la bibliographie (LUQUIENS 1919 [1909], p. 83-84; NYROP 1899, p. 239-240;

TABLE 9.3 – Enclises du pronom personnel masculin objet de troisième personne du singulier.

Seules les configurations dont F > 40 sont données.

Les variations graphiques ne sont pas représentées (p.ex. *qui* correspond à <qui>, <ki> et <qi>.)

Enclise	F
<i>ne l</i>	1471
<i>se l</i>	516
<i>je l</i>	371
<i>qui l</i>	75
<i>que l</i>	42
autres	74

Pronoms personnels 3sg masc. CR atones

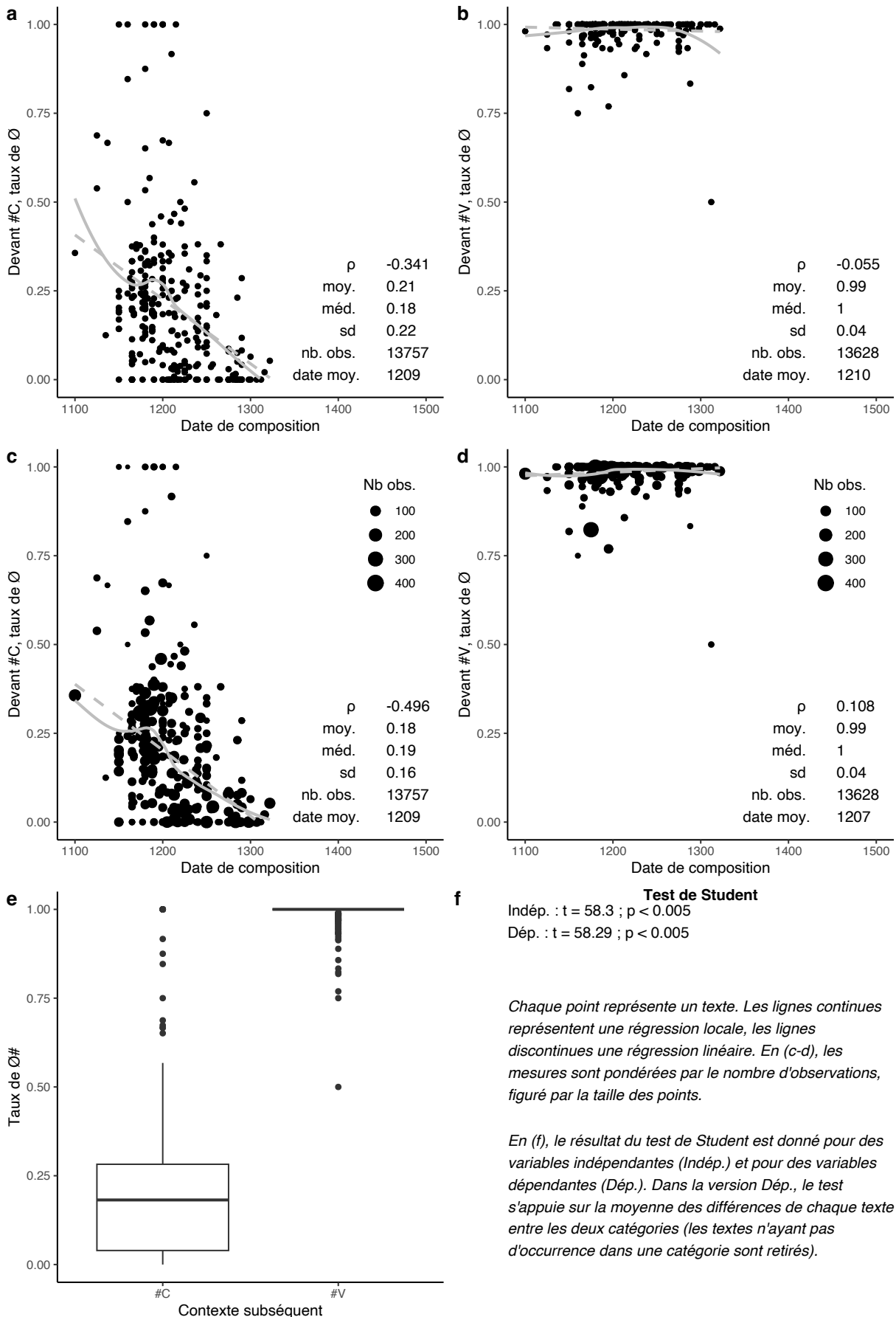


FIGURE 9.1 – Distribution contextuelle des variantes de pronoms personnels de masculin troisième personne du singulier au cas régime en défaut de <e>.

POPE 1966 [1934], p. 323) est clairement visible sur les graphiques de séries temporelles en fig. 9.1a,c : la droite de régression tend clairement vers zéro, et la régression locale ne s'écarte que peu de la régression linéaire. Si le coefficient de corrélation de Pearson est peu satisfaisant sur le jeu de données non pondérées ($\rho = -0,34$), il devient acceptable une fois les taux des textes pondérés par le nombre d'occurrences du pronom dans les textes ($\rho = -0,5$). À moins que les contextes syntaxiques propices à l'enclise ne deviennent par hasard plus rares, cette régression est conforme avec l'idée que le processus ne se maintient que pour les enclises les plus courantes, qui deviennent alors lexicalisées (POPE 1966 [1934], p. 323), et avec celle, plus générale, qui veut que l'enclise disparaisse progressivement dans tous les cas (NYROP 1899, p. 239-240). Il ne faut néanmoins pas oublier que ces données ne portent que sur les enclises représentées, dans le corpus, en deux mots (support d'enclise + *l*) ; or la probabilité que les enclises soient traitées par le corpus comme un seul mot augmente avec le temps, alors que des modifications phonétiques rendent opaque (et probablement lexicalisée) la formation de l'enclise (p.ex. *a+le* est bien représenté *a l* dans le corpus — et non *al*, mais l'évolution en *au* n'est jamais représentée *a u*).

l __#V. Devant initiale vocalique, on ne voit rien de tel : les droites de régression en fig. 9.1b,d sont plates et, si quelques textes s'éloignent de la moyenne (99%), ces écarts tiennent principalement à des textes anglo-normands. Comme illustré en fig. 9.2, le texte, tardif, dont le taux de *l* devant initiale vocalique est le plus faible est anglo-normand, de même que les textes aux taux les plus faibles dans la seconde moitié du xii^e s. De fait, même si l'anglo-normand peut légèrement tirer vers le bas les résultats, l'effet de ce dialecte reste très modéré (a.n. moy. = 95% vs a.fr.cont. moy. = 99%).

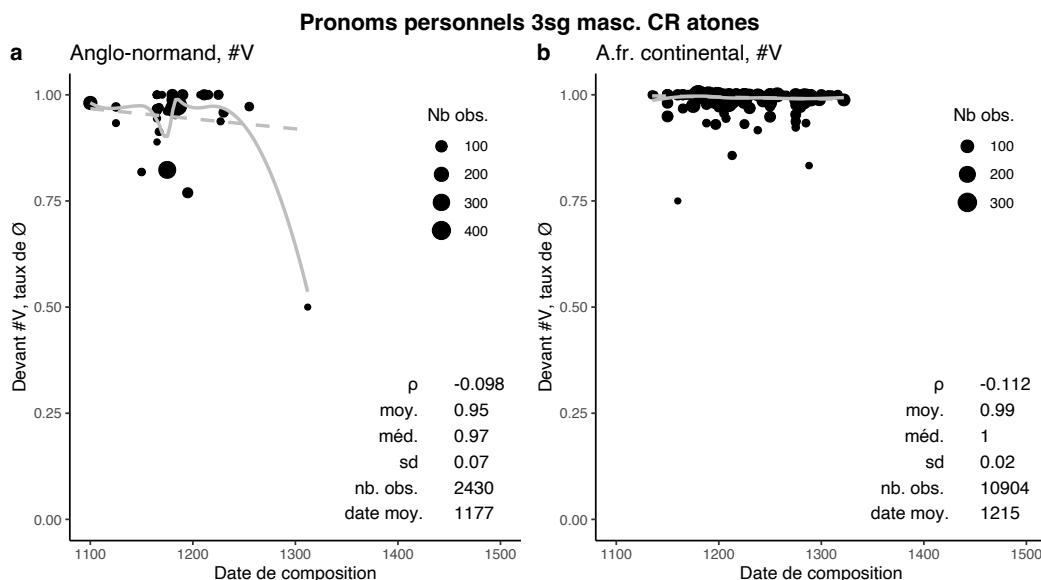


FIGURE 9.2 – Diachronie des défauts de <e>, anglo-normand vs ancien français continental, pronom personnel masculin régime de troisième personne du singulier.

L'effet du contexte, tel que mesuré par le test de student, est donc très fort ($t = 58,3$) et hautement significatif ($p < 0,005$). Comme le représentent les boîtes à moustaches en fig. 9.1 *supra*, cet effet ne stipule pas que le

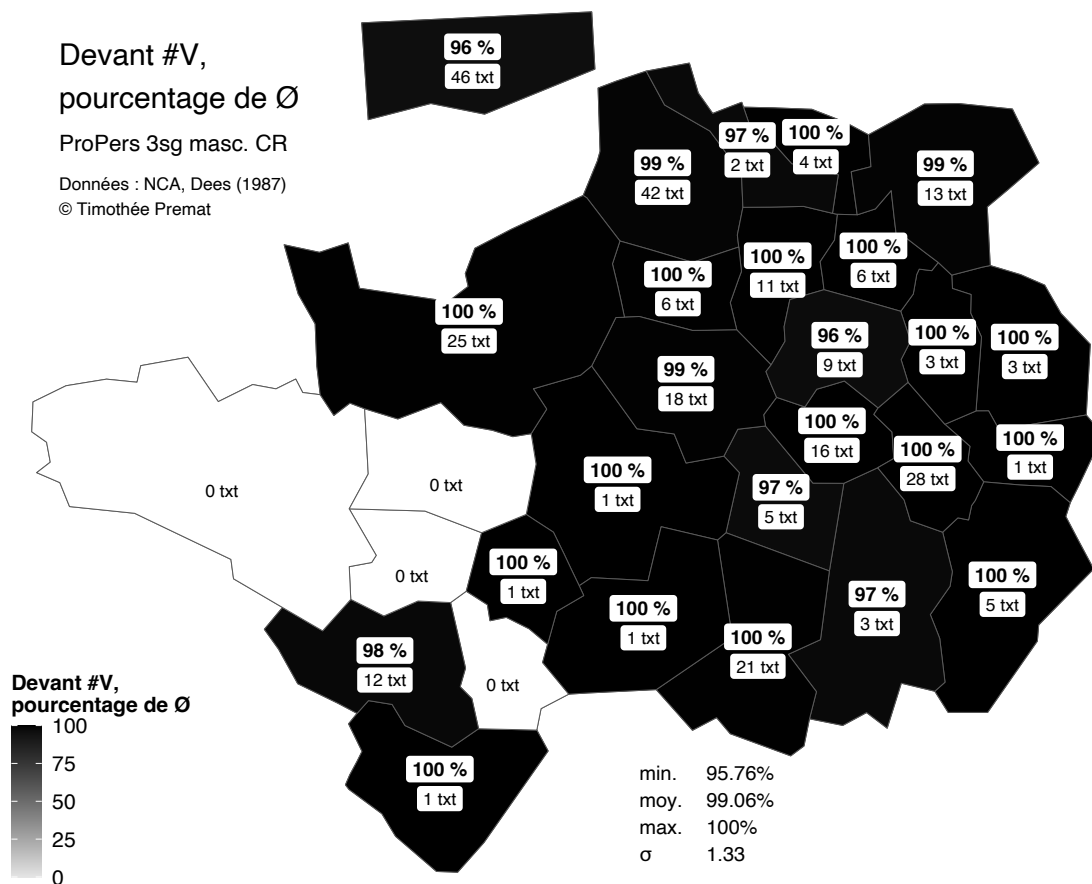


FIGURE 9.3 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, pronoms personnels masculins régimes de troisième personne du singulier suivis d'une initiale vocalique.

contexte prévocalique est une condition nécessaire au défaut de <e> (les cas d'enclise sont nombreux), mais bien que le contexte prévocalique est une condition presque suffisante au défaut de <e>.

9.2.3.2 Distribution diatopique

Comme représenté sur la carte en fig. 9.3, la distribution diatopique des défauts de <e> du pronom personnel objet *le* est très homogène ($\sigma = 1,33$), avec un taux régional minimal de 96% atteint en Angleterre et dans la Marne. Il n'y a donc pas lieu de rechercher une raison diatopique aux quelques cas où *le* n'est pas éliidé devant initiale vocalique. Cette homogénéité étant stable sur l'ensemble de la période, les cartes diachroniques ne sont pas commentées et sont rejetées en annexes (fig. E.8 et E.9, p. 617 - 618).

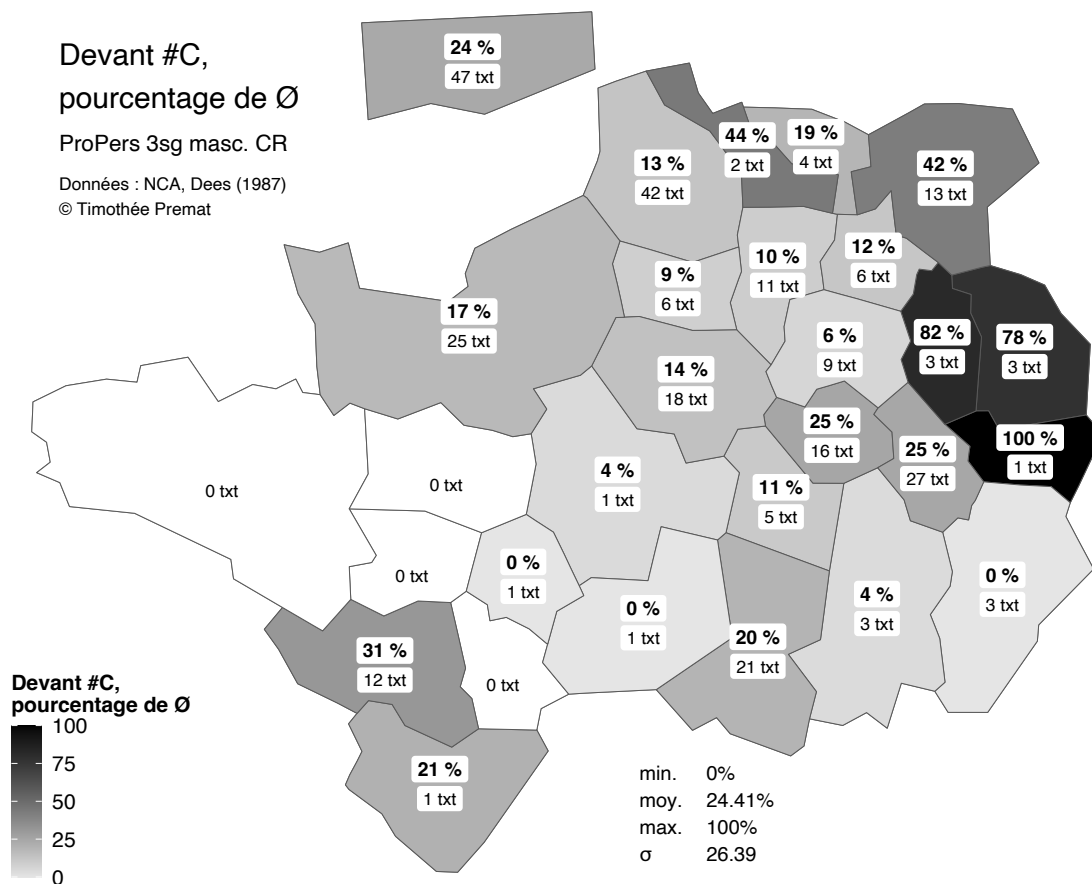


FIGURE 9.4 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, pronoms personnels masculins régimes de troisième personne du singulier suivis d'une initiale consonnantique.

L'enclise, en revanche, semble présenter une certaine affinité pour l'Est et le Nord du domaine, avec des taux de 44% dans le Nord (2 textes), 42% en Wallonie (13 textes), 82% dans la Meuse (3 textes), 78% en région Moselle, Meurthe-et-Moselle (3 textes) et même 100% dans les Vosges (1 texte). Si ces trois derniers taux sont peu consolidés au niveau du nombre de textes, et particulièrement le taux des Vosges (dans le seul texte vosgien, toutes les occurrences du pronom devant initiale vocalique sont en enclise), ce n'est pas le cas de la Wallonie, ce qui permet de légitimer une tendance plus forte à l'enclise dans cette partie du domaine. On peut alors rattacher à cette zone l'Aube (25%, 16 textes) et la Haute-Marne (25%, 27 textes). Néanmoins, les taux ici indiqués sont pleinement dépendants des conditions syntaxiques dans lesquelles les pronoms apparaissent : si des taux élevés témoignent de la vivacité de l'enclise, des taux faibles ne peuvent être interprétés, étant donné qu'ils peuvent être générés par des textes où le pronom n'apparaît pas après de potentiels supports d'enclise.

9.2.3.3 Intersection de la diatopie et de la diachronie

Diachroniquement, cette vivacité de l'enclise dans le Nord-Est est relativement stable pour la Wallonie, seule région à être attestée dans nos quatre périodes ; nous ne faisons ici qu'un rapide commentaire et les cartes sont rejetées en annexes (fig. E.10 et E.11, p. 619 - 620). Les textes des trois régions aux taux maximaux (Meuse, Vosges, région Moselle, Meurthe-et-Moselle) proviennent majoritairement de la seconde moitié du ^{xii} s. (seule la Meuse dispose d'un texte antérieur), mais même à cette époque ces régions disposent des taux les plus élevés. Il n'y a qu'à la fin du corpus, après 1250, que la Wallonie rentre dans le rang (14%, 3 textes), tandis que nos autres régions du Nord-Est à taux élevés ne sont plus attestées.

9.2.3.4 En somme.

Le pronom personnel objet de 3sg masculin connaît des défauts prévo-camiques et préconsonantiques. Les défauts prévo-camiques sont presque systématiques et sont stables sur l'ensemble de la période et sur l'ensemble du domaine ; la spécificité de l'anglo-normand est faible. Les défauts préconsonantiques, qui tiennent à l'enclise avec un terme support antécédent, sont en quantité décroissante, mais cette tendance combine la baisse de la productivité de l'enclise et le fait que Comparalem ne détecte que certaines enclises (celles qui sont écrites en deux mots dans le corpus), et ce alors que l'on suppose que ce sont précisément les enclises que Comparalem ne détecte pas qui sont les plus nombreuses à la fin du corpus.

9.2.4 PROPERSCR2sg : *te* ~ *t*

Le pronom personnel régime de deuxième personne du singulier, dont la forme atone *te* est distincte des formes toniques (*tei*, *toi*, etc.), est sujet à défaut de <e> (formes *t*). Ces formes se réduisant à une consonne apparaissent principalement devant initiale vocalique, les cas d'enclise étant rares. Dans ce qui suit, nous traitons de la distribution diachronique et contextuelle de ces formes *t*, avant d'aborder leur distribution diatopique et l'intersection entre diatopie et diachronie.

9.2.4.1 Distribution contextuelle et diachronique

Les graphiques de séries temporelles en fig. 9.5.a-d, page suivante, représentent l'évolution diachronique des taux de défaut de <e> pour le pronom personnel objet de 2sg *te*.

t __#C. À la différence du pronom régime de 3sg MASC, le pronom régime de 2sg ne connaît pas d'enclise constante : cinq textes disposent d'occurrences préconsonantiques défectueuses, toujours minoritaires, tandis que tous les autres textes n'ont que *te* devant consonne. Ces cinq textes ne disposent que d'une occurrence défectueuse *t* chacun, et c'est le nombre d'occurrences *te* qui compense cette occurrence *t* qui fait varier le taux. Si nous ne sommes pas en mesure de vérifier philologiquement

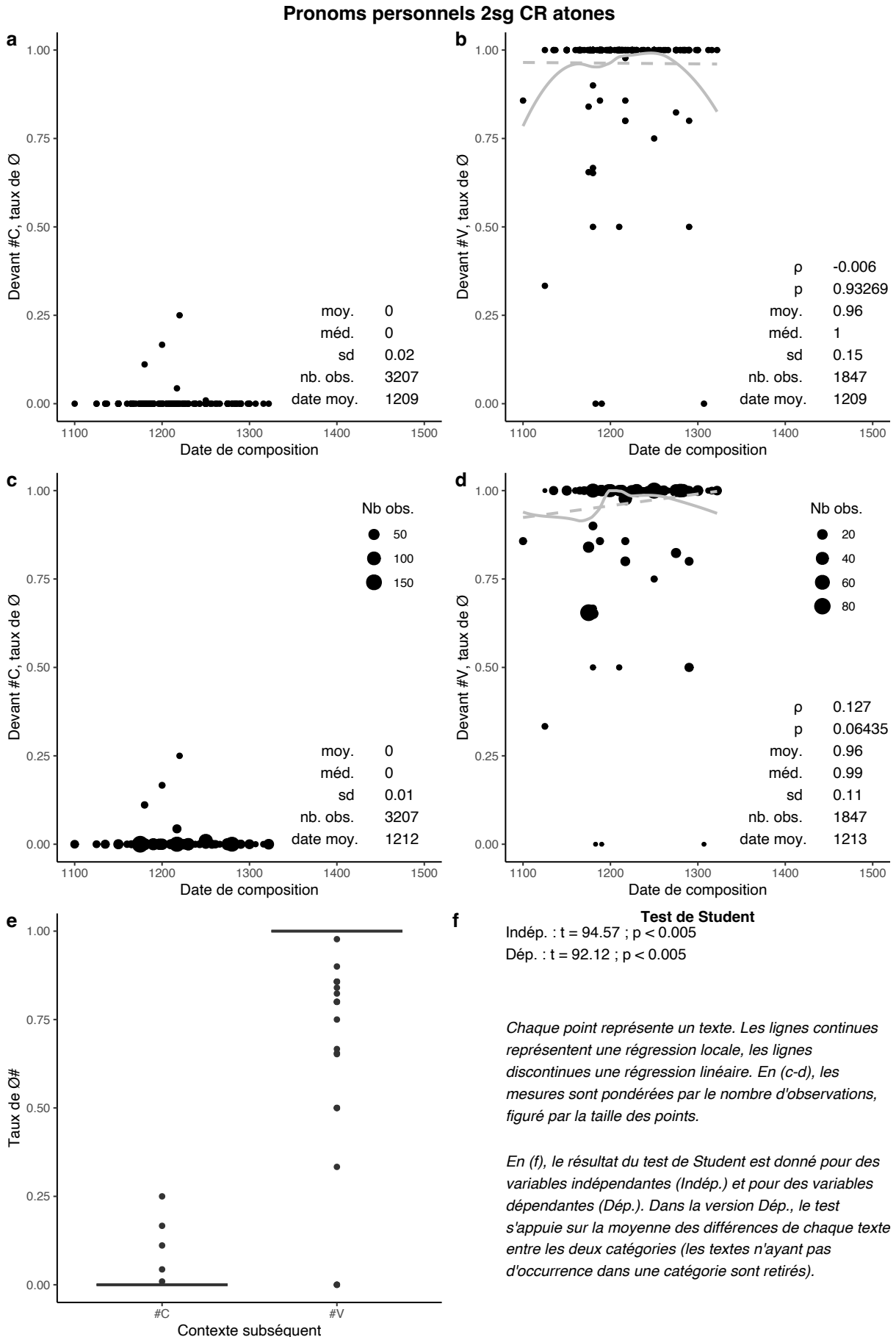


FIGURE 9.5 – Distribution contextuelle des variantes de pronoms personnels de deuxième personne du singulier au cas régime en défaut de <e>.

chacune de ces cinq occurrences, nous notons que 4 occurrences sur 5 sont situées après *ne* ou *je*, qui sont des termes supports d'enclise. Soit ces occurrences sont fautives, soit elles représentent des cas exceptionnels d'enclise de *te* avec d'habituels supports d'enclise. Par ailleurs, dans la cinquième occurrence (9.33), *t* se trouve après *terres*, que WALKER (1981, p. 51) relevait précisément comme un potentiel support d'enclises (*terrel* pour *terre+le*). Néanmoins, en (9.32), le verbe est manquant et n'est pas immédiatement présent dans le contexte syntaxique, aussi il se peut que le passage soit simplement lacunaire et que le verbe manquant soit à initiale vocalique et, en (9.33), *terres* et *t* sont difficilement joignables au sein d'un même domaine syntaxique (on imagine mal comment cela peut être le cas pour le *terrel* de Walker, cité sans contexte). Aussi, si ces occurrences sont tout sauf certaines, il convient néanmoins de noter que *t* devant initiale consonantique n'apparaît qu'après un terme support d'enclise.

- (9.29) *ce que ne t couste rien*
« ce qui ne te coûte rien »
(amo, p. 39, trad. pers.)
- (9.30) *se il autrement ne t creit*
« s'il, pour une autre raison, ne te croit pas »
(perf, p. 12, trad. pers.)
- (9.31) *ie t coniur par dieu lo uiel que quant tu auras fait ton erra t en reuengies por moj*
« je te conjure, par le dieu du ciel, quand tu auras fait ton voyage, reviens pour moi »
(poit, p. 81, trad. pers.)
- (9.32) *je t plus que chevalier que je onques eusse pour la grant proeche et pour la grant loiaute qui est en toi*
« je te [le verbe est manquant] plus que chevalier que j'aie jamais eu, à cause de ta grande prouesse et de ta grande loyauté »
(lancpr, p. 16, trad. pers.)
- (9.33) *je sai bien qui tu es dedens nos terres t molt oi prisier*
« je sais bien qui tu es, dans notre pays j'ai entendu beaucoup de louanges à propos de toi »
(cordres, p. 6, trad. pers.)

amo : JACQUES D'AMIENS (1250*). *L'art d'amors et li remedes d'Amors*. Dees : Nord ; compo. flandr., ca 1250 ; ms. pic. (?), ca 1300 ; MS (JACQUES D'AMIENS 1868) ; DEAF : BestAmOctT.

perf : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS F ; Dees : nil ; compo. champ. mérid., ca 1180 ; ms. Est, ca 1250 ; MS1 (transcr. inédite R. L. H. Lops).

poit : ANON. (1217*). *Chronique traduite de l'Historia Karoli Magni et Rotholandi*. Dees : Vendée, Deux-Sèvres ; compo. saint., ca 1217 ; ms. saint., ca 1238 ; MS1 (AURACHER 1877) ; DEAF : Turpin¹A.

lancpr : ANON. (1220*). *Lancelot*. Ms. S ; Dees : [Somme, Pas-de-Calais] ; compo. nil, ca 1220 ; ms. pic., ca 1316 ; qualité nil (MICHAS 1978-1983) ; DEAF : LancPrM.

cordres : ANON. (1200*). *La prise de Cordres et de Seville*. Dees : Meuse ; compo. lorr., ca 1200 ; ms. Meuse, ca 1262 ; MS2 (DENSUSIANU 1965 [1896]) ; DEAF : PriseCordD.

t __#V. Devant initiale vocalique, la situation est inverse : plus de la moitié des textes ne dispose d'aucune occurrence *te* (en fig. 9.5.b : méd. = 100%), et quelques textes disposent de taux inférieurs à 100%, dont trois n'ont aucun défaut de <e>. Ce sont, au total, 72 occurrences sur 1847 qui ne sont pas en défaut de <e>. Toutes les occurrences sauf une présentent un contexte classique, où *te* précède immédiatement un verbe à initiale vocalique, le plus souvent le verbe « avoir » (30 occ. : 9.34 - 9.36), mais pas uniquement (9.37). Une seule occurrence nous semble suspecte, et ce pour deux raisons : en (9.38), c'est le chiffre 2 qui suit *te*, et c'est donc une initiale consonantique qu'il faut analyser, et non une initiale vocalique ; par ailleurs, le verbe semble manquer dans cet exemple. Il semble donc que, si une occurrence peut être rejetée (dans un texte qui ne contient qu'elle), il faille cependant accepter l'existence de ces contre-exemples. Aussi, le caractère systématique de l'élision de *te* objet devant initiale vocalique ne vaut que pour une majorité des textes, tandis que quelques textes contiennent plus ou moins d'occurrences non élidées.

- (9.34) *mut me repent que te ai crue*
« je regrette beaucoup de t'avoir crue »
(*fablesA*, p. 9, trad. pers.)
- (9.35) *je te auoie comandé la mea chose par doner aus poures por la salu de m arme*
« je t'avais demandé de donner ma possession aux pauvres pour le salut de mon âme »
(*poit*, p. 10, trad. pers.)
- (9.36) *cil qui te aurai en sa possession ne serai vainquz nen esbaiz*
« celui qui t'aurait en sa possession ne serait ni vaincu ni terrifié »
(*pseuturp*, p. 41, trad. pers.)
- (9.37) *se tu viens avec moi et me mostres marsille, je te laisserai aschaper vif; se ce nun, je te occirrai*
« si tu m'accompagnes et me montres Marsile, je te laisserai échapper vivant; sinon, je te tuerai »
(*pseuturp*, p. 39, trad. pers.)
- (9.38) *et nequedain je te yy heres aporter avoques toi lai ou tu dois aler*
« et néanmoins je te [...] deux seigneurs amener avec toi là où tu dois aller »
(*barlaam*, p. 58, trad. pers.)

De ce fait, l'effet du contexte est très fortement marqué, comme figuré par les boîtes à moustache en fig. 9.5.e et par le test de Student en fig. 9.5.f. Cet effet est très fort ($t = 94,57$) et hautement significatif ($p < 0,005$). Si l'on observe la distribution des points en fig. 9.5.e, qui représentent les points extrêmes (*outliers*)³⁰, on observe que la superposition des deux contextes n'a lieu qu'avec les trois points extrêmes dotés d'aucune occurrence *t* devant initiale vocalique; pour tous les autres textes, aucune superposition entre les valeurs des deux contextes n'est possible.

Enfin, la pondération des taux de défaut par texte donne une droite de régression légèrement croissante en fig. 9.5.d, mais dont le coefficient de corrélation ($\rho = 0,13$) et la valeur- p ($p = 0,06$) ne sont pas satisfaisants. En effet, une tendance générale à la systématisation de l'élision devant initiale vocalique ne saurait rendre compte du fait qu'on trouve, jusqu'à la fin du corpus, des textes disposant d'occurrences *te* dans ce contexte. Cet effet est en réalité dû, comme pour les adverbes (sec. 7.3.3), à l'anglo-normand. Comme représenté en fig. 9.6, l'anglo-normand est responsable de taux inférieurs à 100% dont le nombre et le poids sont relativement importants dans la seconde moitié du XII^e s., et ce sont ces textes qui donnent à l'ensemble du domaine une droite de régression pondérée légèrement croissante ($\rho > 0$)³¹, alors que la droite de régression ne concernant que le domaine continental est décroissante ($\rho < 0$ en fig. 9.6.b). Dans tous les cas, que ce soit sur l'ensemble du domaine ou en séparant l'Angleterre et le domaine continental, les coefficients de corrélation sont trop proches de zéro pour être satisfaisants et la valeur- p est trop élevée pour que la mesure soit significative.

fablesA : MARIE DE FRANCE (1180*). *Fables*. Ms. A; Dees : [Angleterre]; compo. NO, ca 1180; ms. a.n., ca 1250; MS (transcription inédite).

poit : ANON. (1217*). *Chronique traduite de l'Historia Karoli Magni et Rotholandi*. Dees : Vendée, Deux-Sèvres; compo. saint., ca 1217; ms. saint., ca 1238; MS1 (AURACHER 1877); DEAF : Turpin^{1A}.

pseuturp : ANON. (1290*). *The burgundian translation of the Pseudo-Turpin Chronicle*. Dees : Haute-Saône; compo. bourg., ca 1290; ms. *idem*; MS (WALPOLE 1949, 1948); DEAF : Turpin^{7W}.

barlaam : ANON. (1217*). *L'histoire de Barlaam et Josaphat*. Dees : Haute-Marne; compo. champ., ca 1217; ms. Est, ca 1290; CR3 (MILLS 1973); DEAF : BalJosPr^{1M}.

30: Ici définis comme les points du premier ou dernier quartile dont la valeur s'écarte de la borne de leur quartile par plus de 1,5 fois l'écart inter-quartile (fonction `geom_boxplot` du package `ggplot2` dans R).

31: En plus d'indiquer à quel point les données correspondent au modèle de la régression linéaire, le coefficient de corrélation de Pearson ρ indique si la corrélation est inverse ($\rho < 0$) ou non ($\rho > 0$).

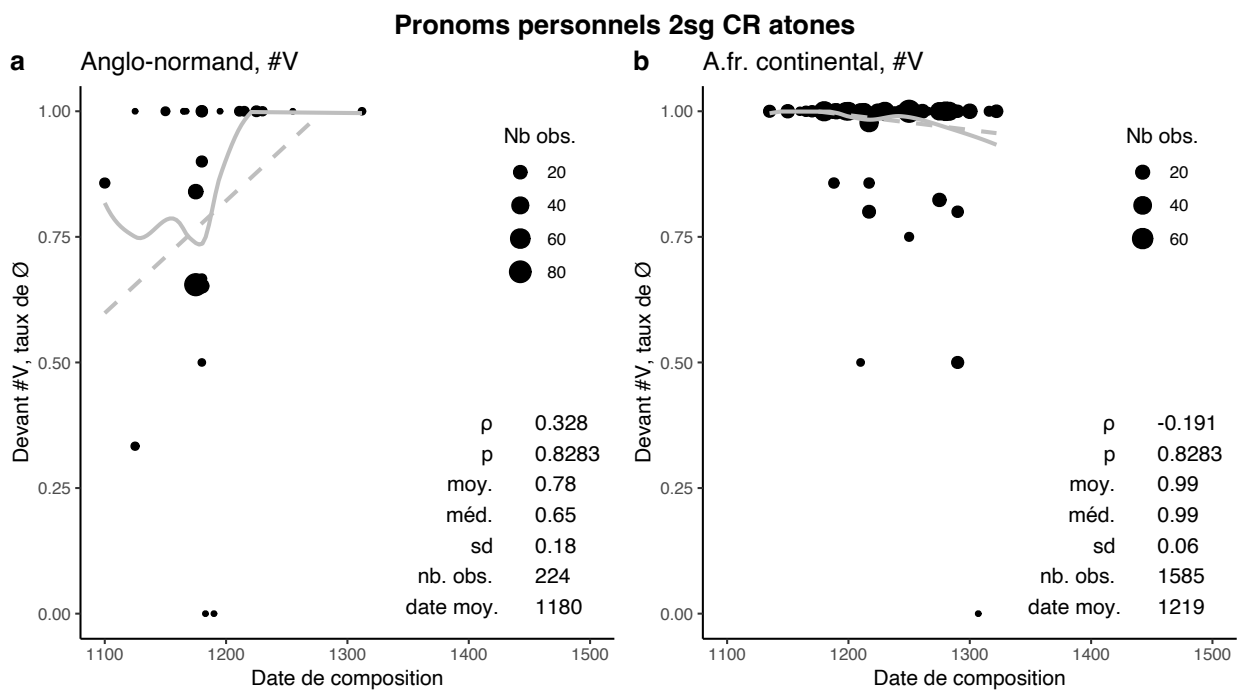


FIGURE 9.6 – Diachronie des défauts de <e>, anglo-normand vs ancien français continental, pronoms personnels régimes de deuxième personne du singulier.

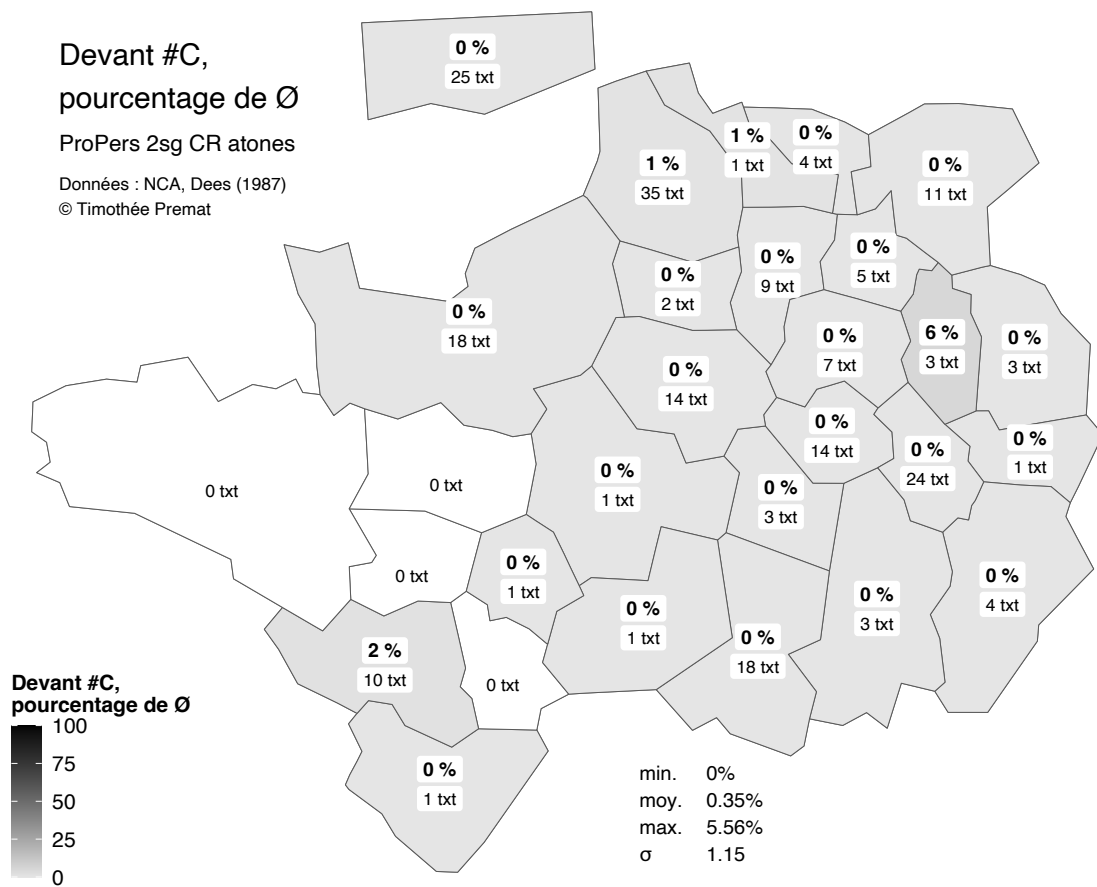


FIGURE 9.7 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, pronoms personnels régimes de deuxième personne du singulier suivis d'une initiale consonnantique.

9.2.4.2 Distribution diatopique

Diatopiquement, le domaine est assez homogène, avec cependant des différences entre les contextes.

t __#C. L'homogénéité est meilleure devant initiale consonnantique que devant initiale vocalique, ce qui est logique étant donnée la très faible dispersion des données préconsonnantiques ($\sigma = 0,02$; σ des taux moyens régionaux = 1,15), comme représenté en fig. 9.7. Les seules régions disposant de taux de défaut de <e> positifs sont celles qui contiennent les textes des 5 occurrences citées en (9.34 - 9.38) *supra*, c'est-à-dire la Meuse (cordres), le Nord (amo), la région Somme, Pas-de-Calais (lancpr) et la région Vendée, Deux-Sèvres (perf, poit), et leurs taux régionaux demeurent toujours faibles, d'une part parce que même ces cinq textes ont plus de *te* que de *t* devant initiale consonnantique, et d'autre part parce que, à part dans le Nord, ils sont compensés par un nombre variable de textes n'ayant aucun défaut devant initiale consonnantique (de 2 textes dans la Meuse à 34 textes en région Somme, Pas-de-Calais).

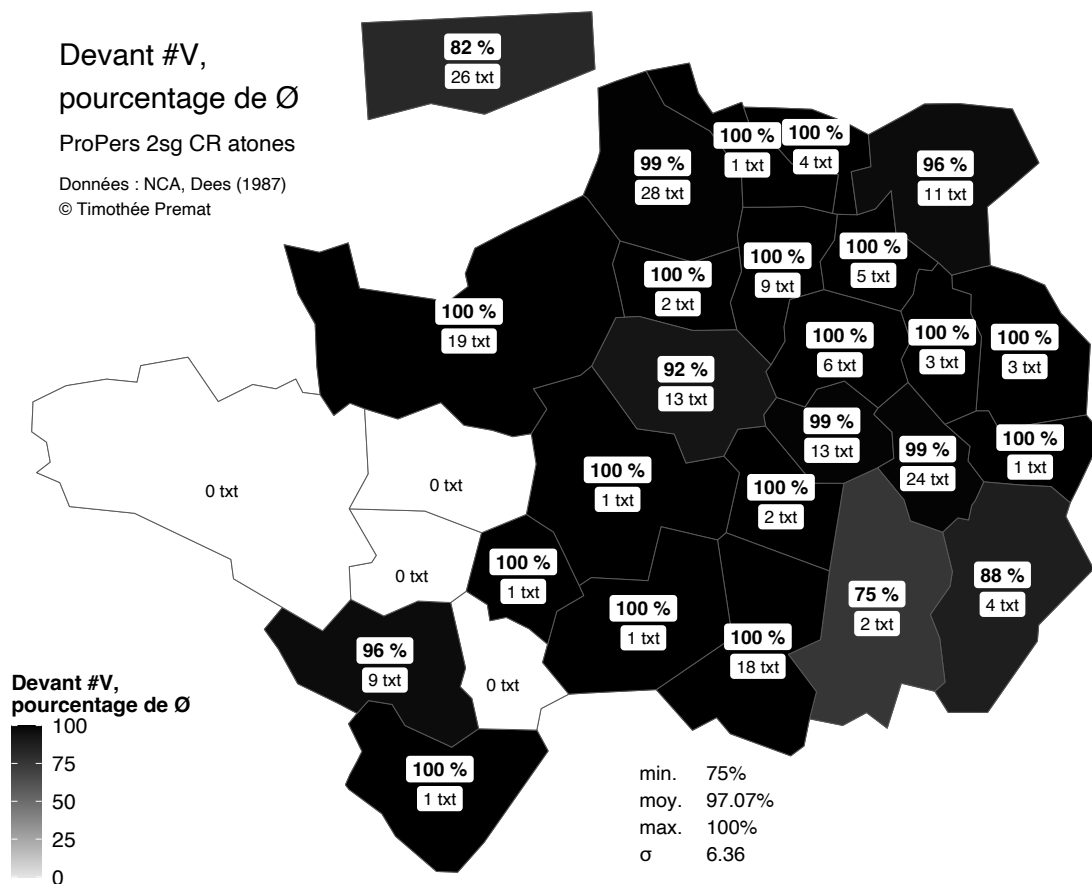


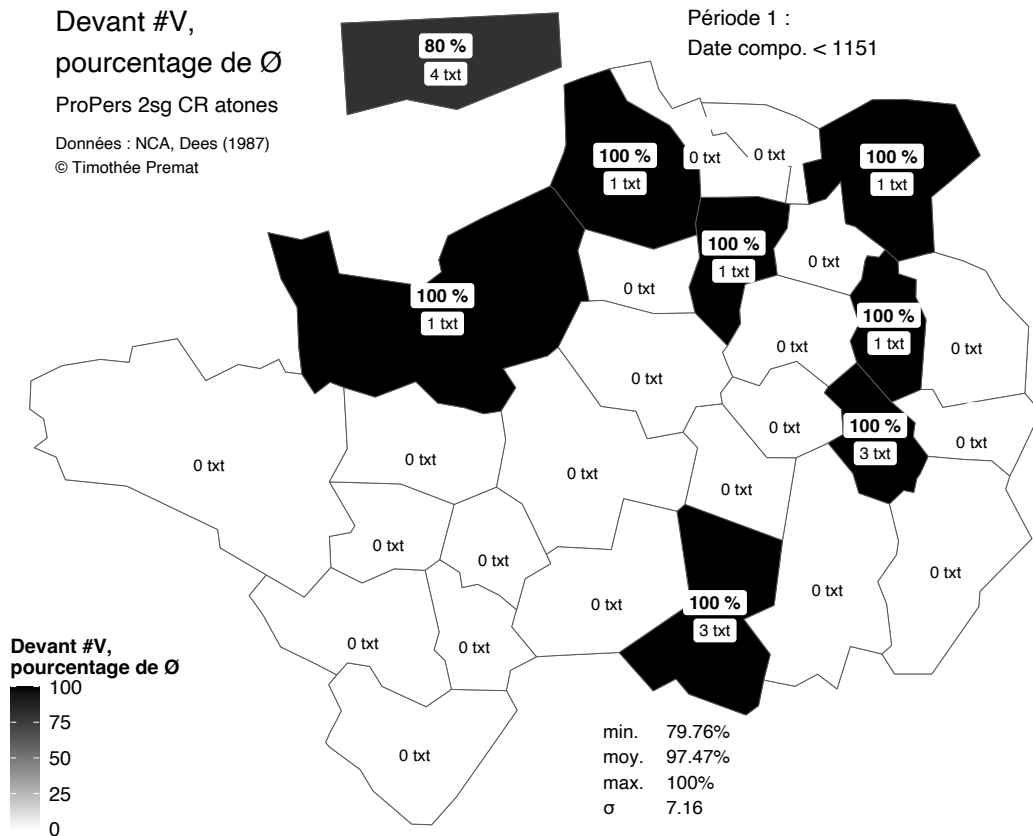
FIGURE 9.8 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, pronoms personnels régimes de deuxième personne du singulier suivis d'une initiale vocalique.

t __#V. Devant initiale vocalique (fig. 9.8), l'homogénéité est moins forte, du moins telle que mesurée par l'écart-type des taux moyens régionaux ($\sigma = 6,36$). De fait, seules trois régions disposent de taux de défaut inférieurs à 90%, la Bourgogne (75%, 2 textes), l'Angleterre (82%, 26 textes) et la Franche-Comté (88%, 4 textes). La région parisienne dispose d'un taux de 92% (13 textes), la Wallonie et la région Vendée, Deux-Sèvres disposent d'un taux de 96% (11 et 9 textes), tandis que toutes les autres régions ont des taux supérieurs à 99%. La présence du taux de 92% en région parisienne s'oppose à tous les autres taux inférieurs à 99%, qui sont tous situés en périphérie du domaine.

9.2.4.3 Intersection de la diatopie et de la diachronie

Le nombre de textes disposant de *t* devant initiale consonantique est si faible que la diachronie de leur diatopie est anecdotique et ne peut renseigner l'analyse générale. Devant initiale vocalique, en revanche, certains éléments peuvent être relevés. Les cartes présentant la distribution diatopique par période sont en fig. 9.9 et 9.10, en pages suivantes.

a



b

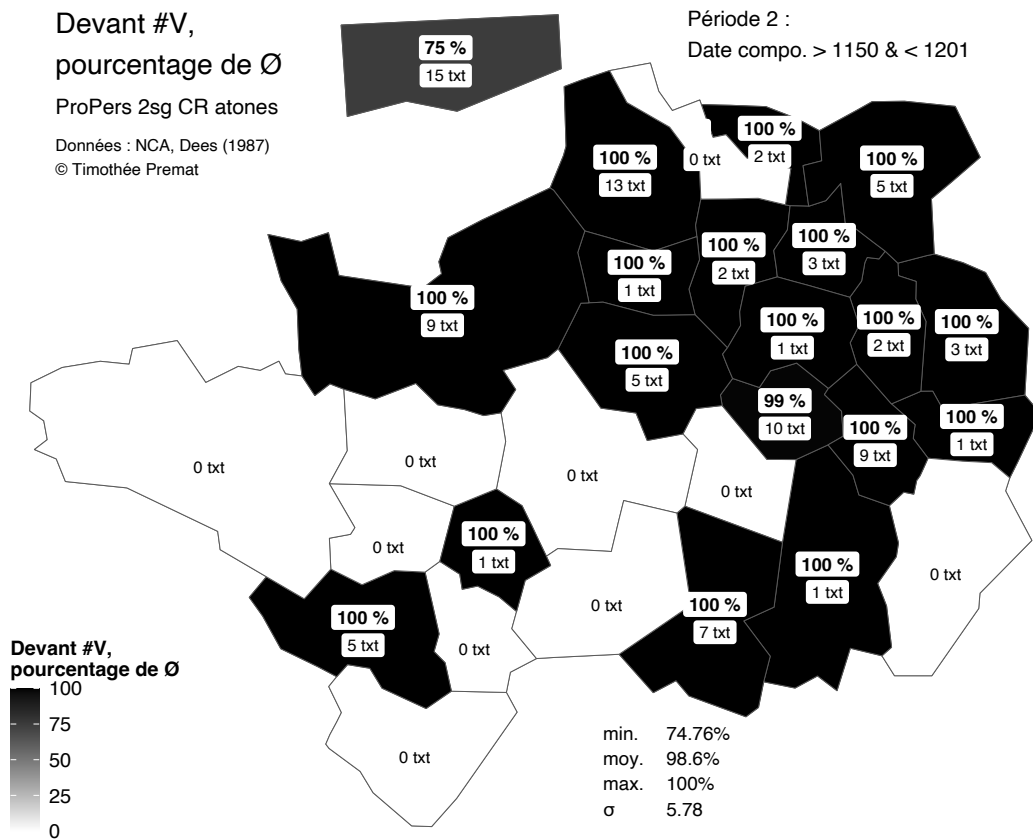
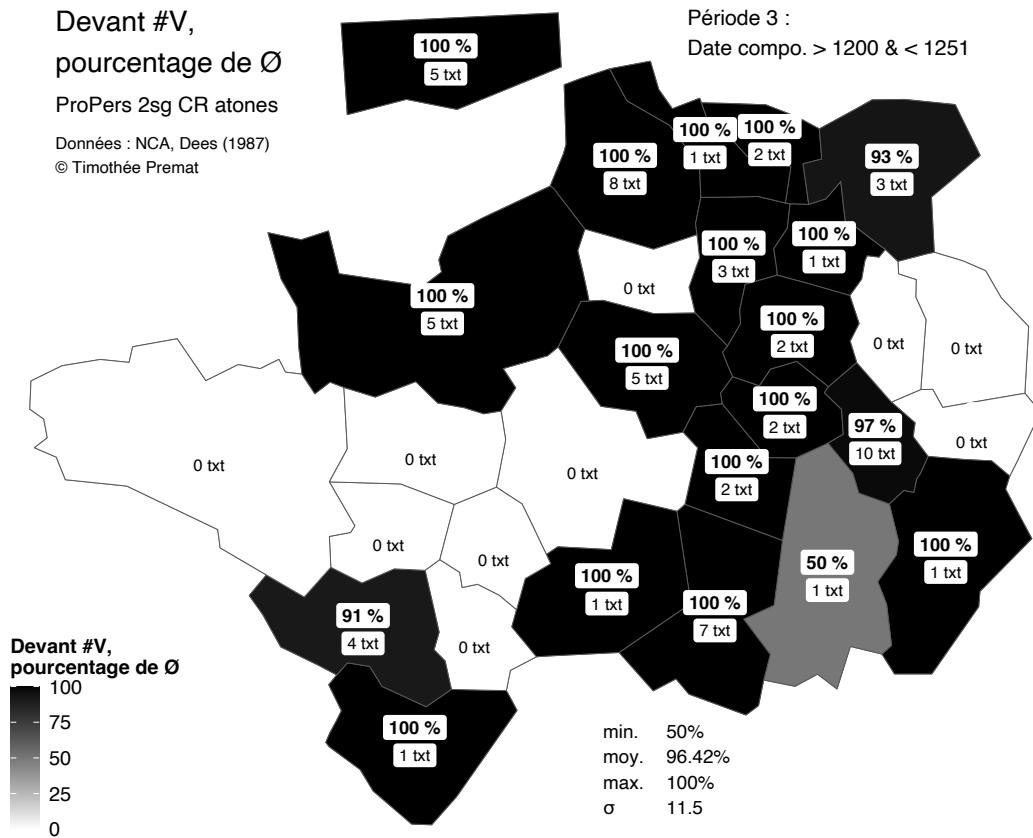


FIGURE 9.9 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1150) et 2 (b : 1151-1200), pronoms personnels régimes de deuxième personne du singulier.

a



b

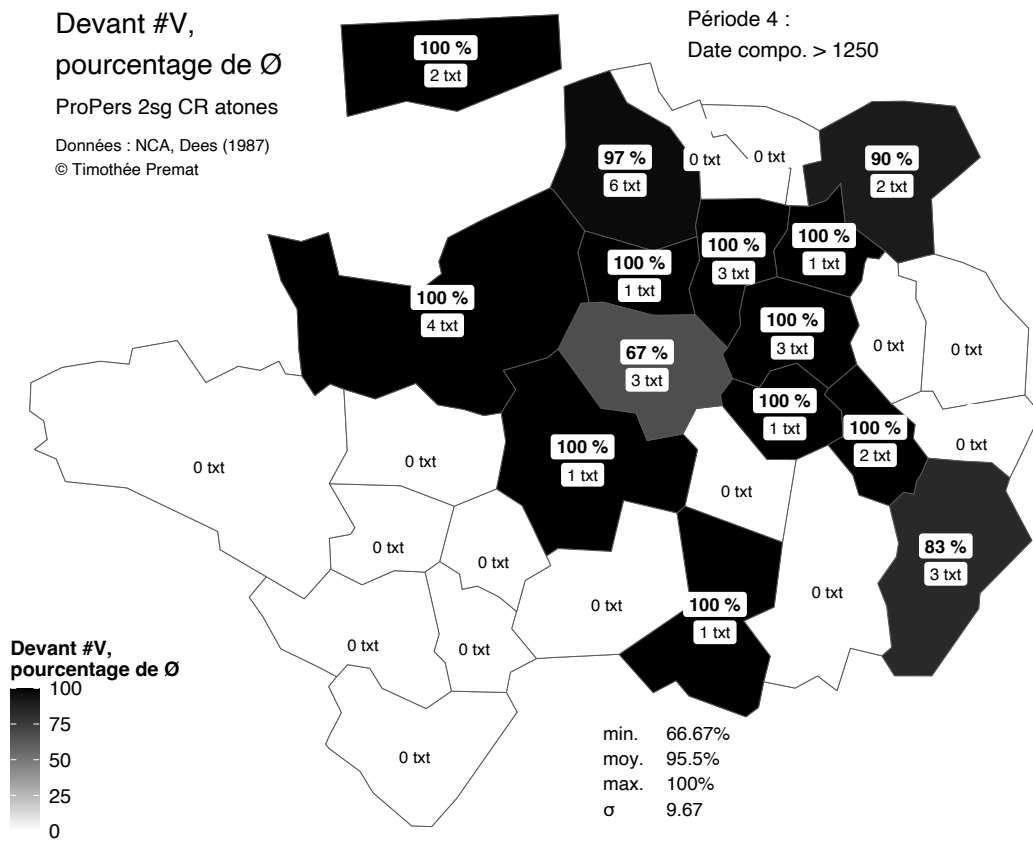


FIGURE 9.10 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), pronoms personnels régimes de deuxième personne du singulier.

Comme cela était déjà visible en fig. 9.6, l'anglo-normand ne dispose de taux spécifiquement bas qu'au XII^e s. Au contraire, les autres régions dont les taux sont inférieurs à 99% n'obtiennent des taux inférieurs à ce seuil qu'au cours du XIII^e s. : la Wallonie est en-deçà de ce seuil sur l'ensemble du siècle (93 et 90%, 3 et 2 textes; fig. 9.10), tandis que la Franche-Comté et la région parisienne ne basculent en-deçà du seuil qu'après 1250, avec 100% de défauts avant 1250 et 83% de défauts en Bourgogne (3 textes) et 67% de défauts en région parisienne (3 textes) entre 1250 et 1322 (fig. 9.10.d). La Bourgogne, non représentée après 1250, bascule également au début du XIII^e s., passant de 100% de défauts (1 texte) dans la seconde moitié du XII^e s. à 50% (1 texte) dans la première moitié du XIII^e s. De même, la région Vendée, Deux-Sèvres passe de 100% (5 textes) dans la seconde moitié du XII^e s. à 91% (1 texte) dans la première moitié du siècle suivant. Aussi, l'intersection entre diatopie et diachronie révèle un effet inverse sur le continent et en Angleterre, avec des taux inférieurs à 99% uniquement au XII^e s. en Angleterre, et uniquement à partir du XIII^e s. sur le continent.

9.2.4.4 En somme

Pour le *PROPersCR2sg te*, les formes défectueuses *t* sont extrêmement rares devant initiale consonantique mais représentent peut-être des cas isolés d'enclise, étant précédées d'un terme support d'enclise. Devant initiale vocalique, les formes *t* sont majoritaires mais ne sont pas hégémoniques; on trouve une certaine quantité de formes *te* qui se maintiennent, au XII^e s. en anglo-normand et au XIII^e s. en Bourgogne, Franche-Comté, Wallonie, région Vendée, Deux-Sèvres et région parisienne. Les autres régions n'ont soit aucune forme *te* devant initiale vocalique, soit un taux de défaut si élevé (< 99%) que leurs occurrences *te* ne sauraient représenter qu'un bruit de fond.

9.2.5 *PROPersCR1sg : me ~ m*

La forme atone du pronom personnel régime de première personne du singulier, *me* (*vs mei, moi*, etc.), connaît, comme le pronom régime de 2sg, quelques cas d'enclise et surtout une domination de l'élision devant initiale vocalique. Dans ce qui suit, nous étudions ces défauts de <e> (formes *m*) en fonction du contexte et de la diachronie, puis en fonction de la diatopie et de l'intersection entre diatopie et diachronie.

9.2.5.1 Distribution contextuelle et diachronique

Les graphiques de séries temporelles en fig. 9.11.a-d, en page suivante, représentent le taux de formes *m* pour le pronom personnel régime de première personne du singulier. La situation décrite par ces graphiques est très proche de celle observée pour les pronoms régimes de deuxième personne du singulier (*t ~ te*), avec une moyenne et une médiane de défauts à 0% devant initiale consonantique, et une moyenne de 97% ou 98% (selon que la moyenne est pondérée par le nombre d'occurrences par texte) et une médiane de 100% devant initiale vocalique. D'emblée, ces simples mesures indiquent que plus de la moitié des textes n'ont que

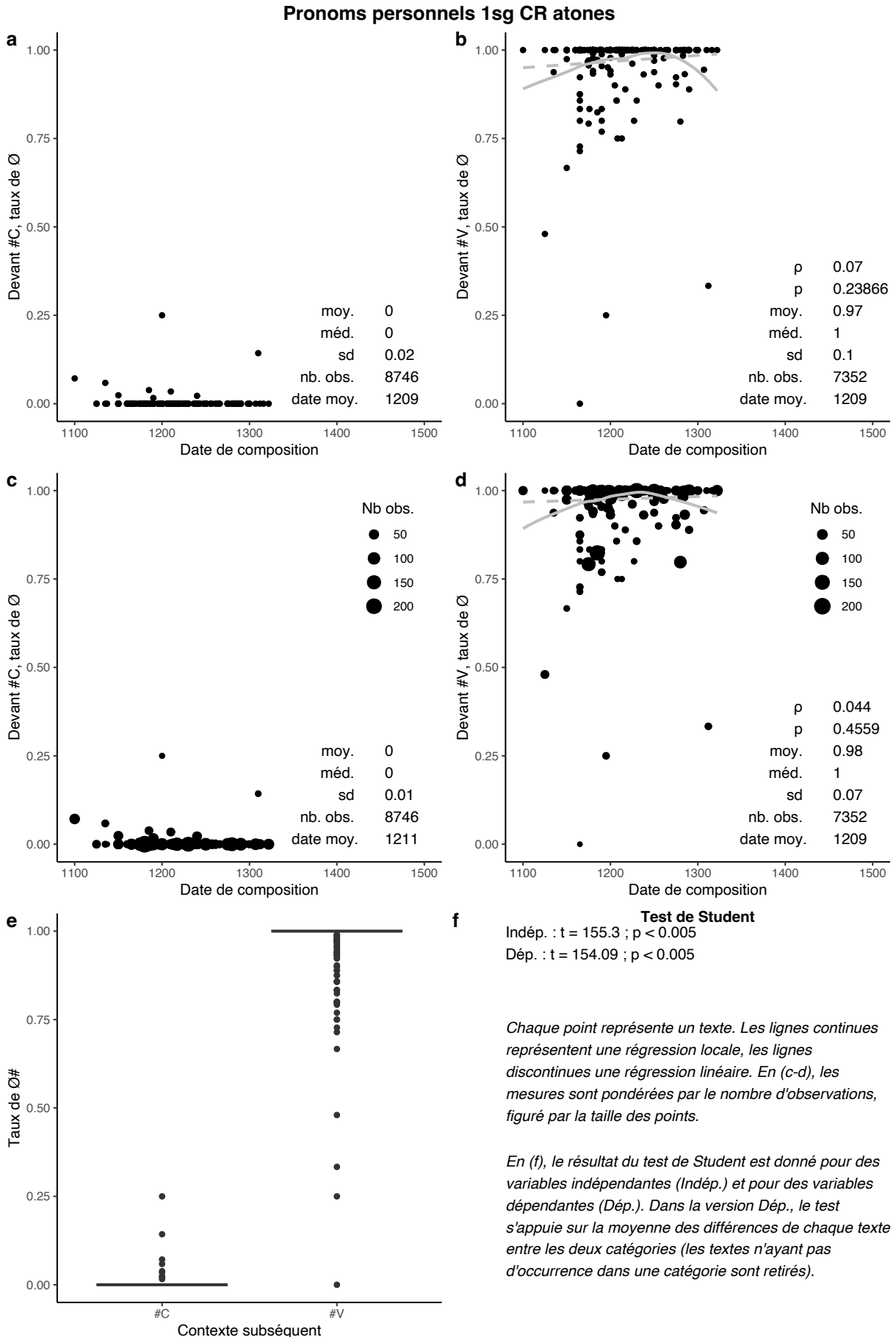


FIGURE 9.11 – Distribution contextuelle des variantes de pronoms personnels de première personne du singulier au cas régime en défaut de <e>.

des formes *me* devant initiale consonantique, et que des formes *m* devant initiale vocalique. Devant initiale vocalique, la régression locale montre une légère hausse des défauts de <e>, y compris sans pondération (ce qui n'était pas le cas pour les pronoms de 2sg CR), mais la conformité des données au modèle linéaire est très faible ($\rho = 0,07$) et non significative ($p = 0,24$). L'effet du contexte, tel que mesuré par le test de Student en fig. 9.11.f, est deux fois plus fort que pour le pronom régime de 2sg (1sg : $t = 155$ vs 2sg $t = 95$), et est hautement significatif ($p < 0,005$). Mais, dans les deux cas, cette valeur- t est à interpréter comme relevant la pertinence d'un effet fort du contexte, et le doublement de la valeur- t pour les pronoms régimes de 1sg n'est associé qu'à une hausse de 1 à 2% de la moyenne des taux de défaut devant initiale vocalique (1sg : 97% ou 98% vs 2sg : 96%).

***m* _#C.** Neuf textes disposent d'occurrences *m* devant initiale consonantiques, à raison d'une occurrence par texte sauf pour *Roland*, qui en contient 4. Dans la moitié des cas, ce *m* est précédé d'un terme support d'enclise et, en particulier pour les occurrences de *Roland*, l'occurrence correspond à une enclise en principe archaïque.

- (9.39) *qui m dona cest chival*
« qui me donna ce cheval »
(*alia*, p. 15, trad. pers.)
- (9.40) *si m depart conme de toi margerite tu m as vencu ta puissance m a confundu*
« et je m'en vais, Marguerite, comme tu m'as vaincu, comme ta puissance m'a confondu! »
(*marga*, p. 8, trad. pers.)
- (9.41) *si m guarisez e de mort e de hunte*³²
« préservez-moi ainsi de la mort et de la honte »
(*Roland*, v. 21, trad. SHORT 1990)
- (9.42) *par voz saveirs se m puez acorder*³³
« si par votre adresse vous me trouvez un accord avec lui »
(*Roland*, v. 74, trad. *id.*)
- (9.43) *si m cumbatrai as [.xii.] cumpaignuns*^{34,35}
« Je combattrai les douze Pairs » »
(*Roland*, v. 878, trad. *id.*)
- (9.44) *ne m fesis mal ne jo ne l te forsfis*³⁶
« tu ne me fis jamais tord, jamais je ne t'en fis »
(*Roland*, v. 2029, trad. *id.*)

alia : ANON. (1185*). *Roman d'Alexandre*. Ms. A; Dees : Vendée, Deux-Sèvres; compo. pic. et occ. orient., ca 1185; ms. poit., ca 1225; MS (ANON. 1937); DEAF : AlexArsL.

marga : WACE (1135*). *La vie de Sainte Marguerite de Wace*. Ms. M; Dees : Haute-Marne; compo. norm., ca 1135; ms. art., 1267; CR3 (WACE 1932a); DEAF : WaceMargaF.

32: BFM : *si me*; Short : *si m'*

33: BFM : *sem*; Short : *se m'*

34: BFM : *sim*; Short : *si m'*

35: L'édition du NCA porte *yy* annoté lemma = "deus" au lieu de l'abréviation pour *douze*.

36: BFM : *nem*; Short : *ne m'*

Pour les quatre exemples de *Roland*, SHORT (1990) édite toujours le terme support, suivi d'une espace, suivie de *m'* (*si m'*, *se m'*, *ne m'*), et MOIGNET (1972a), repris la BFM, édite bien *sem*, *sim* et *nem* (forme agglutinée classique pour rendre l'enclise) pour les trois derniers exemples, mais édite *si me* pour le v. 21, ce qui rend le vers hypermétrique. Sur le ms., représenté en fig. 9.12, on lit bien *S ime*, c'est-à-dire <si me>, dans la mesure où le <m> du copiste ne comporte toujours que trois et non quatre traits verticaux (cf. le <m> de *mort* au milieu du vers) et où l'espace entre l'initiale de vers et le caractère suivant est constant et non significatif dans le ms. Il semble donc, ici, que ce soit l'édition de la BFM qui respecte la source, tandis que l'enclise, par ailleurs sous-entendue par la métrique

du vers, et bien attestée dans les premiers siècles de l'a.fr., est issue d'une intervention éditoriale, en l'occurrence de SEGRE (1971) (pour le NCA) et de SHORT (1990).



FIGURE 9.12 – *Roland*, ms. d'Oxford, fol. 1r, l. 21 (v. 21). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

Quant aux deux autres occurrences précédées d'un terme support traditionnel d'enclise, *qui m* en (9.39) et *si m* en (9.40), elles sont acceptables étant donné que les textes en question datent du XII^e s. Leur acceptabilité est néanmoins légèrement différente : *si m* est bien attesté dans *Roland* et répertorié dans les grammaires (É. BOURCIEZ et J. BOURCIEZ 1974 (1967), p. 45-46; NYROP 1899, p. 239-240; WALKER 1981, p. 51), et le texte est à peine postérieur à *Roland* (compo. ca 1135). En revanche, nous n'avons trouvé aucune attestation de l'enclise *qui m* dans les grammaires consultées (qui sont certes avares en description du paradigme des enclises), mais celles-ci attestent bien l'enclise analogue *quil*; si la forme *quil* est à retenir comme une enclise et non comme une simple élision (ce que semble indiquer l'enclise rare de *qui+les* → *quis*; Joly), alors elle indique que *qui* est un terme potentiellement support d'enclise, mais que celle-ci est en principe restreinte au pronom de 3sg CR. Par ailleurs, NYROP (1899, p. 239-240) atteste également de l'enclise *ki+se* → *kis* dans *Roland*, ce qui semble indiquer que l'enclise avec *qui* en terme support n'est pas strictement limitée à *le*. Dans ce cas, *qui m* est acceptable comme occurrence rare d'une enclise *qui+me*, certes à la fin du XII^e s. (alia, compo. ca 1185). Les six autres occurrences de *m* devant initiale consonantique apparaissent après un mot qui n'est pas habituellement support d'enclise et sont donc suspectes, et nous les laissons de côté.

m __#V. Devant initiale vocalique (fig. 9.11.b,d), 177 formes conservent leur <e>. Une écrasante majorité de formes *me* devant voyelle se trouve devant un verbe, et vingt formes se trouvent devant une préposition (*me en*, *me a*) ou une conjonction (*me e*). Ces deux derniers cas indiquent que le pronom, tout en ayant la forme atone, occupe une position en principe dévolue à la forme tonique, comme illustré en (9.45 - 9.47).

- (9.45) *aidiés me a redrechier je me veul acorder aidiés me a rapayer*
« aidez-moi à faire amende, je veux me réconcilier, aidez-moi à rembourser [ma dette] »
(pritheo, p. 1, trad. pers.)
- (9.46) *por fol uenistes hui enprendre contre me en champ de bataille*
« vous êtes venu follement vous en prendre à moi sur le champ de bataille »
(pap, p. 29, trad. pers.)
- (9.47) *sustien me e sals serai*
« aide-moi et je serai sauvé »
(psautier, p. 107, trad. pers.)

pritheo : ANON. (1290*). *Li priere Theophilus*. Dees : Somme, Pas-de-Calais; compo. pic., ca 1290; ms. pic., ca 1300; CR2 (SCHELER 1877); DEAF : PriereTheopS.

pap : ANON. (1175*). *La vie du pape Saint Gregoire*. Ms. A1; Dees : nil; compo. Nord, ca 1175; ms. tour., ca 1238; MS1 (SOL 1977) DEAF : SGreg_a¹S.

psautier : ANON. (1125*). *Livre des psaumes*. Dees nil; compo. a.n., ca 1125; ms. a.n., ca 1150; CR2 (MICHEL 1876), DEAF : PsCambrM.

FOULET (1924). « L'accent tonique et l'ordre des mots, formes faibles du pronom personnel après le verbe ».

Cet usage de formes atones en position tonique n'est pas isolé, mais il demeure ici marginal, et peut indiquer que, malgré sa forme faible, le pronom est tonique (FOULET 1924, p. 66). Ici, il ne fait guère de doute que

l'absence d'élision de *me* procède du fait que *me* et le terme suivant à initiale vocalique n'appartiennent pas au même constituant syntaxique, la préposition ou la conjonction introduisant un autre constituant. En revanche, dans tous les autres cas, où *me* précède immédiatement un verbe à initiale vocalique, *me* semble bien être l'objet du verbe, et il n'est ici pas possible de supposer une frontière syntaxique entre les deux éléments qui soit à même de neutraliser l'élision, comme en (9.48 - 9.50).

- (9.48) *tant comme je avoye plus prouffité en philosophie et en theologie de tant me esloignaye plus des philosophes et des services divins par ordure de vie*
« plus j'étais instruit en philosophie et en théologie plus je m'éloignais des philosophes et des services religieux à cause de la débauche »
(abe, p. 10, trad. pers.)
- (9.49) *cist se purpense en son quer dist 'me est avis dame*
« celui-ci réfléchit dans son cœur et dit : « À mon avis, dame, ... » »
(edmond, p. 4, trad. pers.)
- (9.50) *ore me est avis qe*
« à mon avis... »
(ipo, p. 20, trad. pers.)

Aussi, si 20 occurrences sur 177 peuvent se justifier par la position forte du pronom de forme faible, suivi d'une frontière syntaxique, ce n'est pas le cas pour les 157 autres occurrences, et il faut donc admettre que le caractère très fortement dominant de *m* devant initiale vocalique n'est pas hégémonique et laisse passer quelques occurrences *me*, en l'occurrence entre 2 et 3% des occurrences.

abe : JEAN DE MEUN (1288*). *Traduction de la première épître de P. Abélard*. Dees : région parisienne ; compo. frc., ca 1280 ; ms. Paris, ca 1395 ; MS3 (JEAN DE MEUN 1934).

edmond : MATTHIEU PARIS (1255*). *La vie de saint Edmond*. Dees : nil ; compo. a.n., ca 1255 ; ms. a.n., ca 1280 ; MS2 (J.-A. BAKER 1929) ; DEAF : SEdmCantB.

ipo : HUE DE ROTELANDE (ca 1185). *Ipomedon*. Compo. a.n., ca 1185 ; ms. a.n., ca 1190 ; CR (transcription inédite).

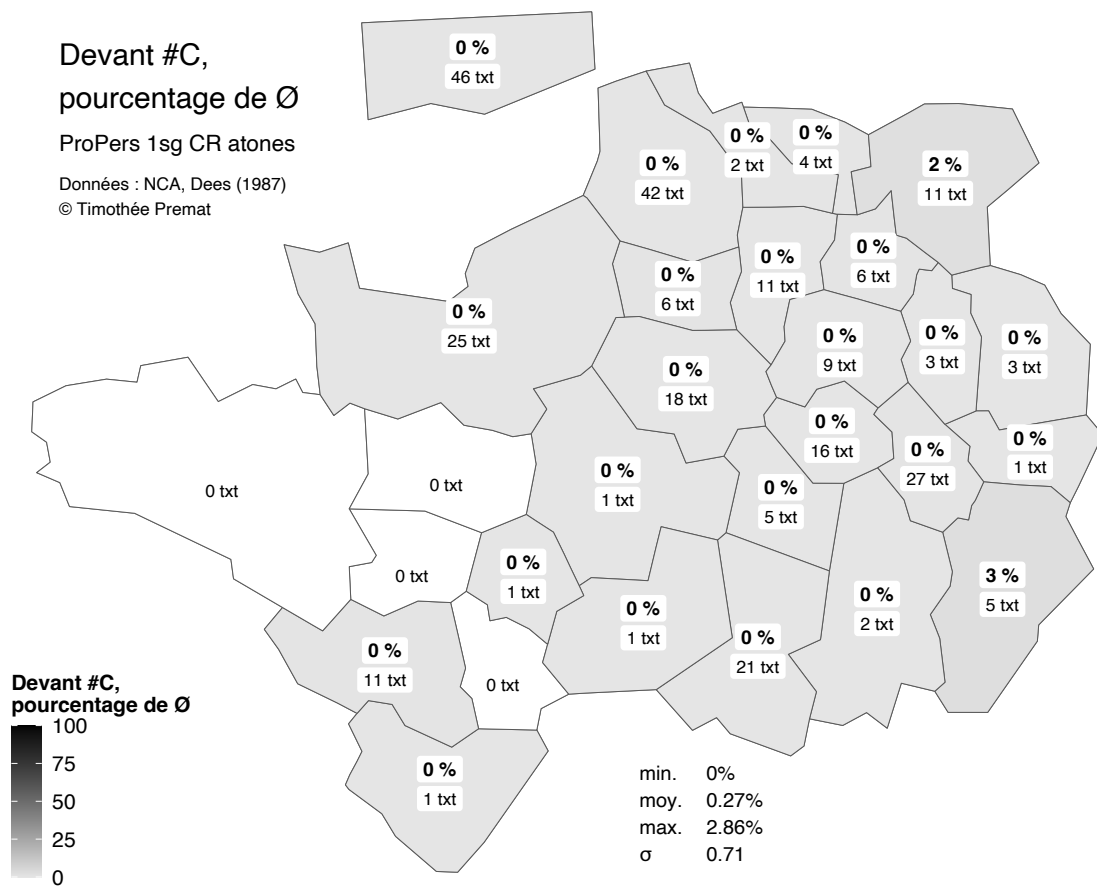


FIGURE 9.13 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, pronoms personnels régimes de première personne du singulier suivis d'une initiale consonnantique.

9.2.5.2 Distribution diatopique

m __#C. Du fait du faible nombre de textes (9 textes) disposant de formes *m* devant initiale consonnantique (12 occ.), dont nous avons montré *supra* qu'il s'agissait de cas d'enclise, avec certitude pour la moitié des occurrences et avec plus de doutes pour l'autre moitié, l'ensemble du territoire est doté d'un fort degré d'homogénéité, jamais atteint dans les parties du discours et/ou lemmes étudiés jusqu'à présent ($\sigma = 0,71$). Comme représenté en fig. 9.13, certains textes comportant des occurrences *m* disparaissent de la carte par l'effet de l'arrondi. Ainsi, les 4 occurrences de *Roland* sont noyées dans les 1180 occurrences *me* des autres textes anglo-normands, et le taux régional est arrondi à 0%. De même, les isolées (une seule occurrence par texte, un seul texte de ce type par région) disparaissent dès que les régions comportent un nombre satisfaisant de textes ; c'est le cas en région Vendée, Deux-Sèvres avec *a lia* dont l'occurrence *m* en (9.39) disparaît dans l'arrondi sur 11 textes, ainsi que dans la Meuse, où l'occurrence *m* de *marga* en (9.40) disparaît dans l'arrondi sur 27 textes. Seules deux régions ont des taux moyens supérieurs à 0,5%, à savoir la Franche-Comté (3%, 5 textes) et la Wallonie

Vendée, Deux-Sèvres) correspondant, pour le pronom *me*, à des taux compris entre 98 et 100%.

Pour le pronom régime de 1sg, cette singularité de l'Angleterre est due à deux caractéristiques. Comme illustré en fig. 9.15, d'une part le domaine anglo-normand dispose à la fois de moins de textes ne présentant aucune occurrence *m* devant initiale vocalique que le domaine continental (en termes relatifs), ayant à la place de nombreux textes dont les taux se situent dans la fourchette de 70-99%, qui correspond peu ou prou à la fourchette dans laquelle le français continental connaît ses taux les plus bas (> 75%). D'autre part, l'anglo-normand dispose aussi de taux dont les valeurs sont extrêmement basses, avec notamment un texte qui ne connaît que des *me* devant initiale vocalique et trois textes dont les taux de *m* devant initiale vocalique sont compris entre 25 et 50%. De tels taux n'existent pas sur le continent, et tirent vers le bas le taux anglo-normand.

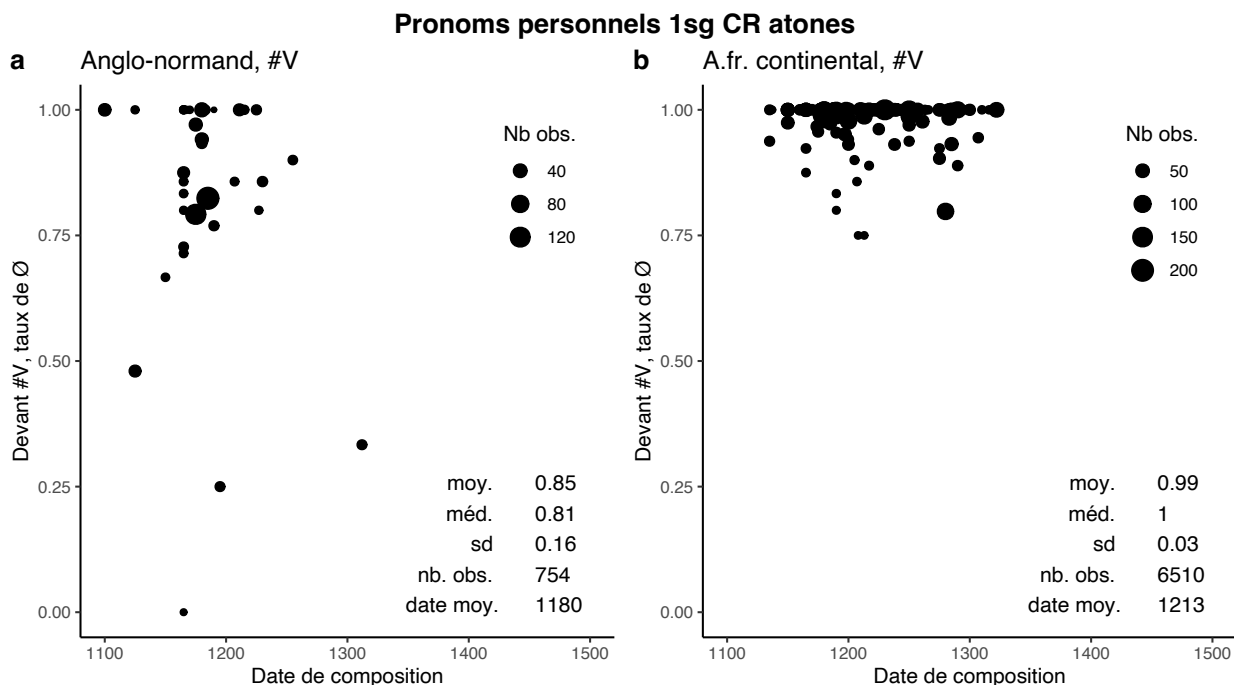


FIGURE 9.15 – Diachronie des défauts de <e>, anglo-normand *vs* ancien français continental, pronoms personnels régimes de première personne du singulier.

Il en résulte une différence, faible ($t = -3,55$) mais hautement significative ($p < 0,005$), entre l'ensemble anglo-normand et l'ensemble continental, telle qu'illustrée en 9.16, en page suivante. Sur les boîtes à moustaches de cette figure, on observe que la zone de variation commune relevée *supra* en a.n. et a.fr.cont. (entre 70 ou 75% et 100%) ne concerne pas la même quantité relative des données : en anglo-normand, cette zone représente le premier et une partie du second quartile des données, alors qu'en a.fr.cont., elle n'est couverte que par des points à valeurs extrêmes (*outliers*, c'est-à-dire des points du premier quartile considérés comme statistiquement marginaux, leur valeur s'écartant de la frontière supérieure de leur quartile par plus de 1,5 fois l'écart inter-quartile). De ce fait, la moitié inférieure de la variation non marginale de l'a.n. correspond à la variation marginale de l'ensemble continental, tandis que la variation marginale de l'a.n. s'étend sur des valeurs que l'a.fr.cont. ne connaît tout

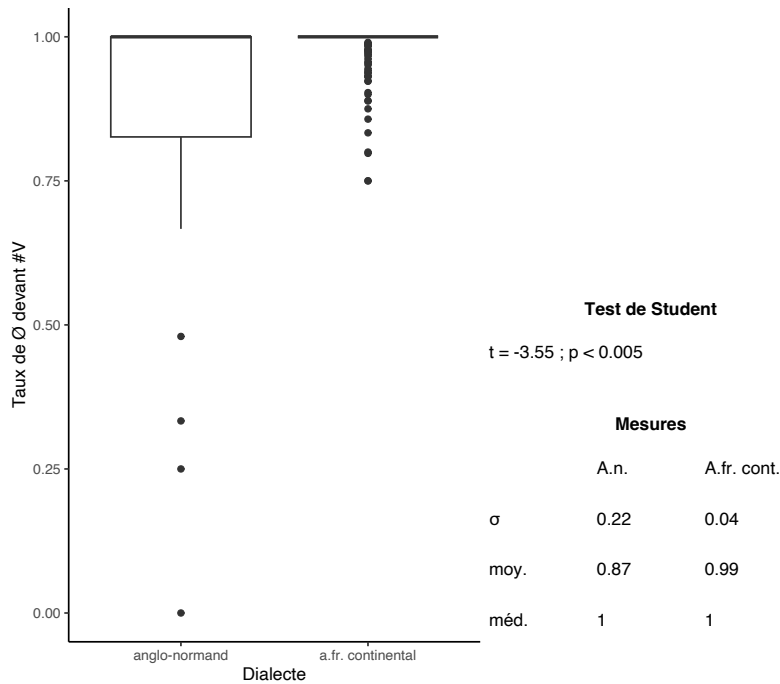


FIGURE 9.16 – Diachronie des taux de défaut de <e>, anglo-normand vs ancien français continental, pronoms personnels régimes de première personne du singulier devant initiale vocalique.

simplement pas. Néanmoins, l'effet est faible du fait que plus de 50% des données sont concentrées à 100% pour l'a.n. comme pour l'a.fr.cont. (méd. = 100% dans les deux cas). Corollairement, l'anglo-normand est bien moins homogène que l'ensemble continental (a.n. : $\sigma = 0,22$ vs a.fr.cont. : $\sigma = 0,04$), une situation que nous avons observée pour les adjectifs féminins singuliers (sec. 5.3.5) et les prépositions (sec. 6.3.5), mais non pour les adverbes (sec. 7.3.5). L'anglo-normand se singularise donc non pas par des taux plus faibles en général, mais par la co-existence de textes où *me* est toujours *m* devant initiale vocalique, et par l'existence de nombreux textes où *me* n'est pas toujours, rarement ou jamais élidé en *m* devant initiale vocalique, tandis que, sur le continent, les textes où *me* n'est pas toujours élidé en *m* sont marginaux, et les formes *m* dominant toujours largement même au sein de ces textes.

9.2.5.3 Intersection de la diatopie et de la diachronie

m _#C. Devant initiale consonantique, les cartes diachroniques ne font que reproduire les observations que nous avons proposées *supra* sur les occurrences en (9.39 - 9.44) et, étant donné le faible nombre d'occurrences en question (12 occ.), il n'est pas approprié de tenter d'en déduire une répartition diatopique en tant que telle. Les cartes diachroniques en question sont rejetées en annexes, en fig. E.12 et E.13 (p. 621 - 622).

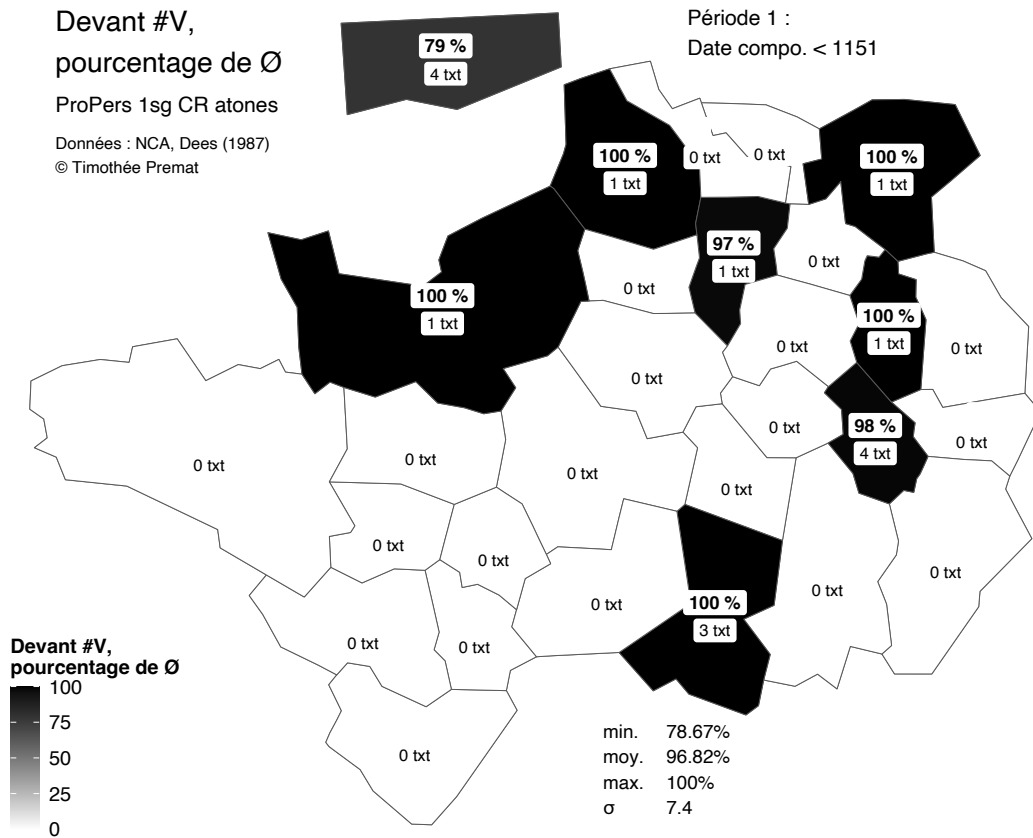
m __#V. Devant initiale vocalique, le découpage des cartes par périodes ne montre pas que la distribution diatopique serait plus hétérogène à un moment donné qu'elle ne l'est en synchronie, comme représenté sur les cartes 9.17 et 9.18 en pages suivantes. Certes, l'écart-type des taux régionaux est substantiellement plus élevé que l'écart-type synchronique ($\sigma = 3,6$) en périodes 1 (1100-1050 : $\sigma = 7,4$) et 4 (1251-1322 : $\sigma = 10,17$), mais cela est principalement dû au plus faible nombre de régions à taux élevés (98-100%) attestées à ces périodes, qui ne nivelle pas vers le bas l'écart-type, tandis que l'observation manuelle de la distribution des taux confirme la bonne homogénéité du domaine, Angleterre mise à part.

L'Angleterre, qui dispose du taux minimal en synchronie (87% en fig. 9.14), dispose des taux minimaux dans la première moitié du *xii*^e s. (79%; fig. 9.17.a) et au *xiii*^e s. (92% dans la première moitié du siècle, 62% dans la seconde; fig. 9.18). Dans la seconde moitié du *xii*^e s., seule la région Moselle, Meurthe-et-Moselle dispose d'un taux inférieur à celui de l'Angleterre, et cette différence est minime (Angleterre : 89%, 34 textes *vs* Moselle, Meurthe-et-Moselle : 88%, 3 textes). Le caractère marginal en synchronie du domaine anglo-normand, relativement à la situation continentale, est donc récurrent, mais la force de cette marginalité varie : l'anglo-normand se rapproche le plus du taux continental dans la première moitié du *xiii*^e s. (a.n. : 92% *vs* ensemble du domaine : 99%), alors que c'est aux extrémités du corpus que l'anglo-normand diffère le plus de la situation de l'ensemble du domaine, avec 79% dans la première moitié du *xii*^e s. (*vs* 97% pour l'ensemble du domaine) et 62% après 1250 (*vs* 96% pour l'ensemble du domaine).

9.2.5.4 En somme

Devant initiale consonantique, les occurrences *m* témoignent de rares cas d'enclise du pronom régime de 1sg; *me* est la règle dans ce contexte hors enclise. Devant initiale vocalique, le domaine continental est caractérisé par une variation marginale entre 70 et 99% de *m*, tandis que cette variation est moins marginale et plus ample en anglo-normand. Aussi, s'il faut admettre l'existence d'occurrences *me* devant initiale vocalique, celles-ci sont rares sur le domaine continental, et ne sont jamais dominantes au sein d'un texte, alors qu'elles sont moins rares en anglo-normand. Mais, même en anglo-normand, l'élision de *me* devant initiale vocalique est la situation de loin la plus courante (taux moyen de *m* : 87%). Cette distribution est relativement stable dans le temps, l'anglo-normand étant toujours le premier ou le second dialecte disposant du moins de *m* devant initiale vocalique. On peut donc établir que l'élision est la règle, que cette règle n'est pas respectée pleinement dans une proportion marginale de textes continentaux, et dans une proportion non marginale mais minoritaire de textes anglo-normands.

a



b

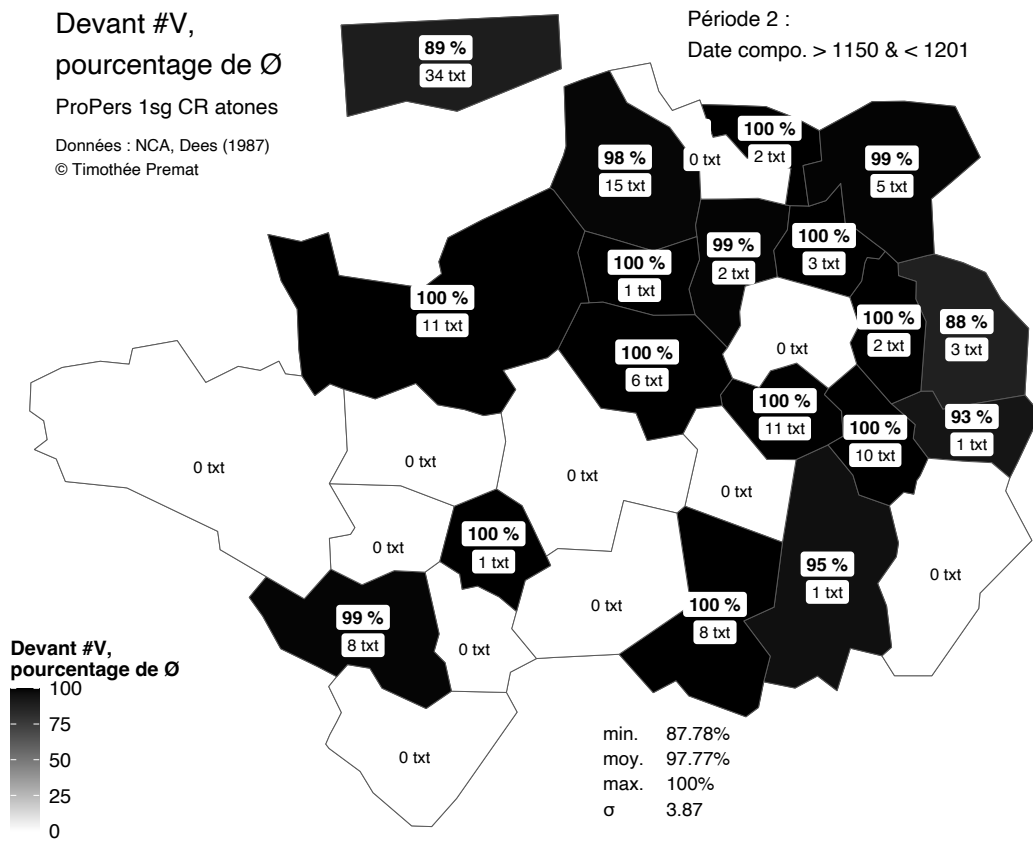
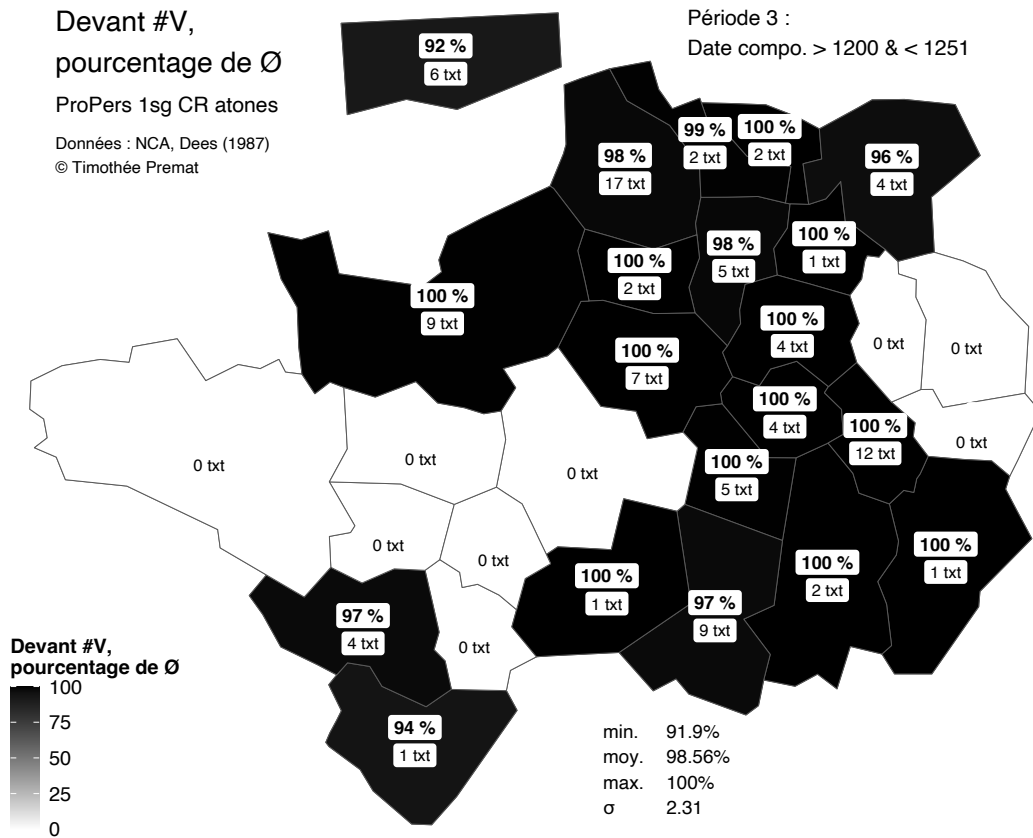


FIGURE 9.17 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1050) et 2 (b : 1151-1200), pronoms personnels régimes de première personne du singulier, devant initiale vocalique.

a



b

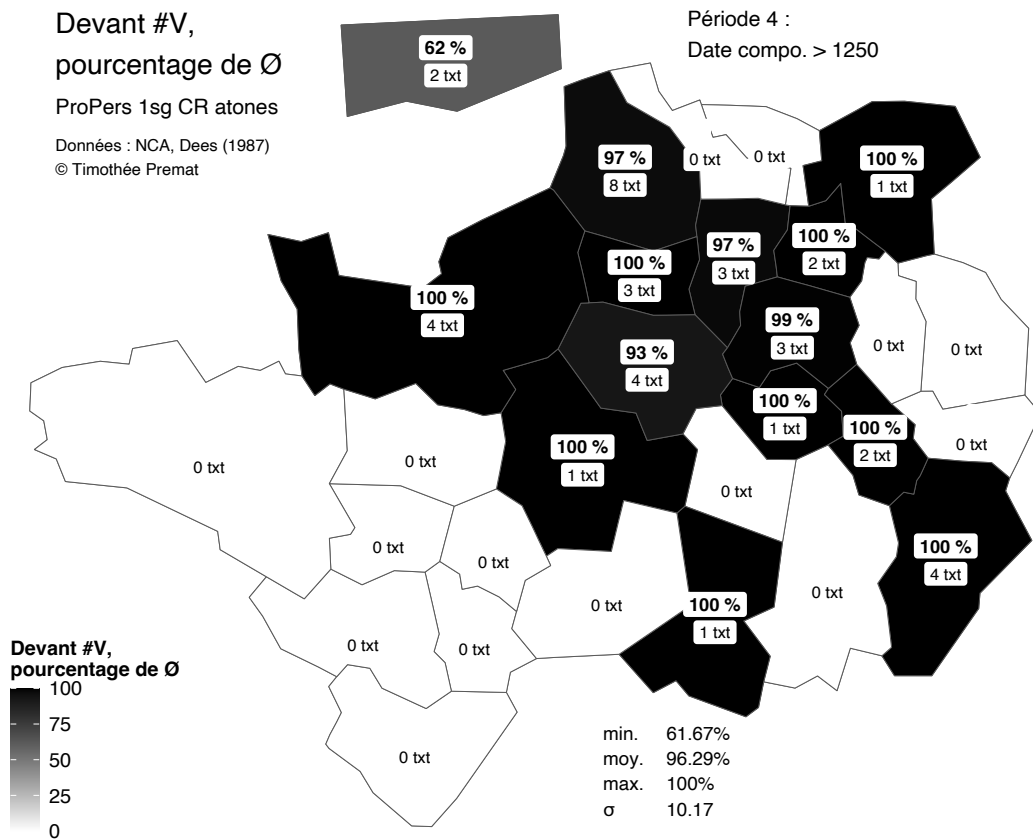


FIGURE 9.18 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), pronoms personnels régimes de première personne du singulier, devant initiale vocalique.

9.2.6 PROPERSCS1sg : *je, jou* (etc.) ~ *j*

Le pronom personnel sujet de première personne du singulier ne connaît pas de division stricte entre les séries toniques et atones. Comme exposé en sec. 9.1.1.1, il apparaît d'abord sous la forme d'une unique série (formes *je, jo, jeo, jou*), employées tant pour les usages toniques que atones, et susceptibles d'élisions métriques dès *Roland* et graphiques dès le *Lapidaire* de Philippe de Thaon (compo. 1^{er} t. XII^e s.), puis deux séries semblent se séparer mais elles demeurent homographes, à l'exception de la forme *gié*, qui est toujours tonique et apparaît au XII^e s. (MARCHELLO-NIZIA 2020c, GGHF, p. 680). Dans ce qui suit, nous traitons de la distribution contextuelle et diachronique des formes défectueuses (formes *j, g, i*), avant de traiter de leur distribution diatopique et de l'intersection entre leurs distributions diatopique et diachronique.

Ici, comme expliqué en sec. 9.1.2 et 9.2.2 *supra*, nous n'avons pas isolé les formes atones, mais avons pu exclure les formes toniques non homographes d'une forme atone et n'appartenant pas à la période où le pronom est indéterminé du point de vue accentuel (c'est-à-dire les formes *gie, gei, iei* etc.); nous avons également exclu toutes les formes annotées comme CS mais ayant la forme d'un CR (107 occurrences du type *me, moi, m*, etc.).

9.2.6.1 Distribution contextuelle et diachronique

Les graphiques de séries temporelles en fig. 9.19.a-d, page suivante, représentent l'évolution diachronique des taux de défaut de voyelle finale atone des pronoms personnels sujet de première personne du singulier. Cette exclusion seulement partielle des formes toniques a pour conséquence directe un taux moyen de formes défectueuses (type *j : g, i, j*) devant initiale vocalique bien inférieur à celui observé pour les pronoms régimes *supra*, de 48 ou 46% (selon que la pondération en fonction du nombre d'occurrences par texte n'est pas appliquée ou l'est; fig. 9.19.b,d). De fait, se focaliser uniquement sur les formes en C<e> (*je, ge*) ne permet pas d'exclure les formes toniques, *je* étant une forme appartenant aux trois séries : à la série indéterminée archaïque, à la série atone et à la série tonique (cf. table 9.1 *supra*); au total, sur les 3192 formes prévocaliques du pronom retenues en 9.19, 1229 sont des formes non défectueuses du type *je*.

Malgré les limitations induites par cette impossibilité de séparer formes toniques et atones, les graphiques en fig. 9.19a-e montrent qu'une différence fondamentale entre les deux contextes se maintient. Cette différence est moins marquée ($t = 21,57$) qu'elle ne l'était pour les pronoms dont les formes toniques et atones pouvaient être séparées (3sg MASC CR : $t = 58$; ³⁷ 2sg CR : $t = 95$; 1sg CR : $t = 155$ ³⁸) mais demeure hautement significative ($p < 0,005$) et d'une force convenable, comparable à celle de la préposition *entre* ($t = 23$)³⁹ et sans commune mesure avec celle des adjectifs féminins singuliers ($t = 5$)⁴⁰ ou des adverbes ($t = -3$).⁴¹ Comme pour *entre*, cette puissance de l'effet traduit une condition nécessaire et non suffisante : le pronom doit être placé devant initiale vocalique pour pouvoir perdre sa voyelle.

37: Cf. fig. 9.1.

38: Cf. fig. 9.5 et 9.11.

39: Cf. fig. 6.17.

40: Cf. fig. 5.10.

41: Cf. fig. 7.8

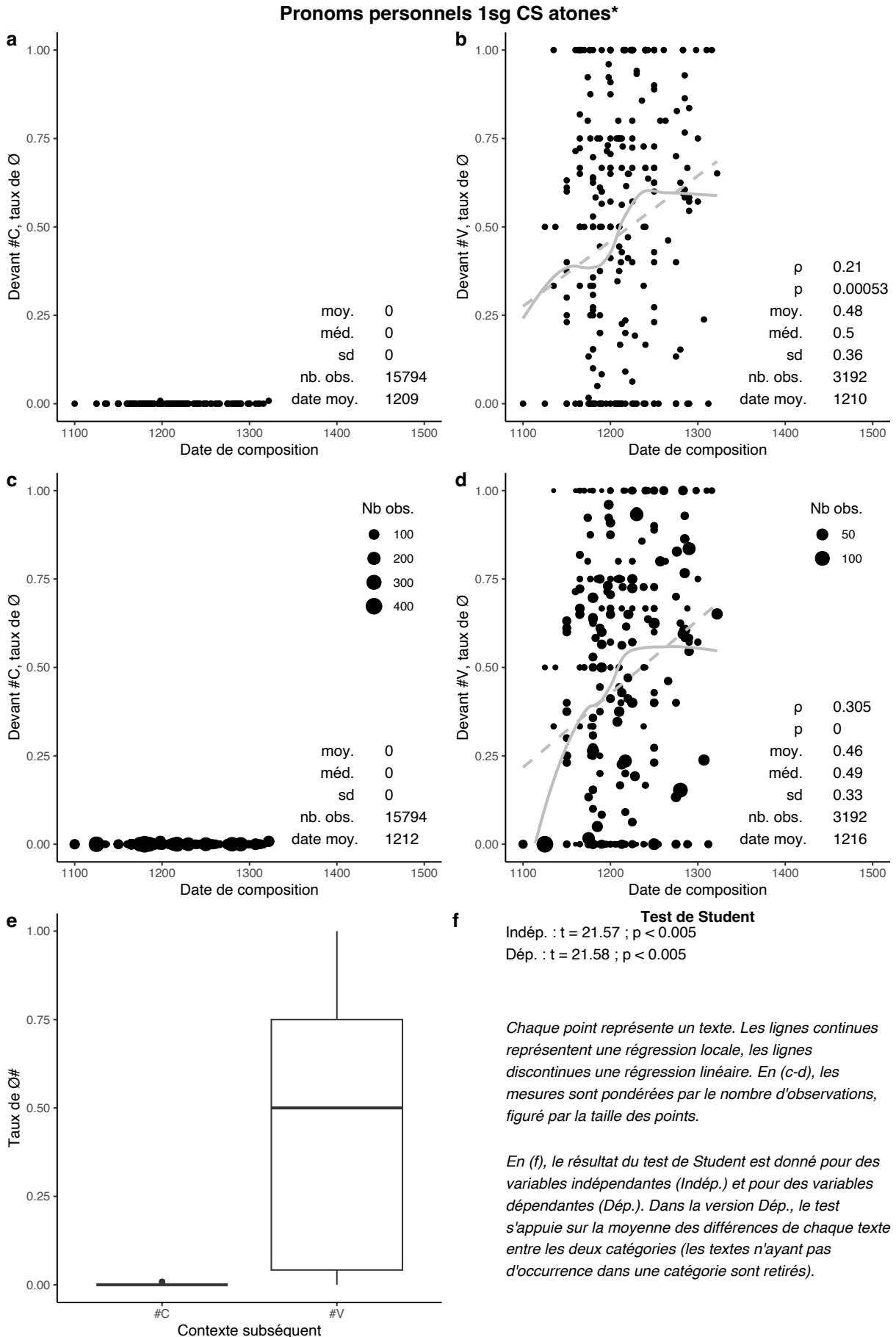


FIGURE 9.19 – Distribution contextuelle des variantes de pronoms personnels de première personne du singulier au cas sujet en défaut de <e>.

*Atones : formes atones, formes indéterminées ou homographes.

j __#C. Devant initiale consonantique, Comparalem n'a détecté que trois occurrences *i* ou *j* annotées comme étant des PROPERSONNELS. Ces occurrences sont douteuses et, étant donné leur quantité par trop marginale (3 sur 15 794 occ.), nous les ignorons et considérons d'emblée que le pronom personnel sujet de 1sg n'est jamais défectueux devant initiale vocalique.

Une telle conclusion n'est possible qu'en l'absence d'enclise de *je*; si *je* peut être support d'enclise (*jo+le* → *jol*, *jo+les* → *jos*, *jes*)⁴², aucune des grammaires que nous avons consultées ne fait état d'une enclise dont *je* soit, non le support, mais l'enclitique. De fait, dans les contextes où *je* est enclitique en fr.mod., c'est-à-dire dans les inversions sujet-verbe (*que vois-je?*), en a.fr. c'est une forme non défectueuse qui est employée, indéterminée au niveau accentuel le plus souvent (9.51), spécifiquement tonique parfois (9.53); pour les pronoms dont le statut atone ou tonique n'est pas ambigu, la règle est que ce soit la forme tonique qui est postposée si elle termine le groupe (MÉNARD 1988 [1973], p. 57). Le contexte de l'inversion sujet-verbe est, au demeurant, fondamentalement différent des contextes d'enclise vus jusqu'à présent, dans la mesure où le terme support est, en fr.mod., un verbe, et non un mot fonctionnel monosyllabique comme c'est d'ordinaire le cas en a.fr.

- (9.51) *por ce cuit ie qu il iut a li*
« à cause de cela, je crois qu'il fut gisant, alité »
(percevalb, p. 78, trad. pers.)
- (9.52) *de ce me puis ie bien uanter*
« de cela, je peux bien me vanter »
(perl, p. 9, trad. pers.)
- (9.53) *que uostre mere vos ait appris ne enseingnié de ce mie ne uos blaim gie*
« que votre mère ne vous ai rien appris ni enseigné à ce propos, je ne vous en blâme pas »
(perm, p. 34, trad. pers.)

j __#V. Devant initiale vocalique, Comparalem a détecté 3192 occurrences du pronom personnel sujet de 1sg. Le taux moyen de défaut par texte s'établit à 48%, ou 46% après application d'une pondération des textes en fonction de leur nombre d'occurrences. De nombreux textes se répartissent aux deux extrémités de l'échelle des ordonnées, n'ayant que des formes défectueuses ou que des formes non défectueuses devant initiale vocalique, mais la plupart des textes (cf. boîtes à moustaches en fig. 9.19.e) disposent d'un taux intermédiaire. Si le taux de chaque texte, pris individuellement, n'a pas grande signification étant donné l'absence de discrimination entre formes atones (élidables) et toniques (non élidables), la tendance d'ensemble est remarquable : on trouve de plus en plus de formes élidées. Cette tendance, telle que figurée par la droite de régression en fig. 9.19.b,d, n'est pas satisfaisante du point de vue d'une corrélation linéaire ($\rho = 0,21$ ou $0,3$), ce qui est naturel étant donnée la distribution élatée des données, mais elle est néanmoins significative ($p < 0,005$)⁴³, ce qui indique que l'absence de corrélation entre la date de composition et les taux de défaut peut être rejetée.

Cette tendance peut s'interpréter de deux façons différentes, qui peuvent être additionnées. D'une part, il est possible que l'élosion des pronoms atones soit de plus en plus forte. D'autre part, il est possible que les

42: LA CHAUSSÉE 1989 [1974], p. 120-121; NYROP 1899, p. 240; POPE 1966 [1934], p. 323.

percevalb : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS B; Dees : Haute-Marne; compo. champ. mérid., ca 1180; ms. bourg. sept., ca 1310; MS1 (transcr. inédite R. L. H. Lops); Dupliqué : perb.

perl : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS L; Dees : Aube; compo. champ. mérid., ca 1180; ms. champ. mérid., ca 1275; MS1 (transcr. inédite R. L. H. Lops).

perm : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS M; Dees : Nièvre, Allier; compo. champ. mérid., ca 1180; ms. bourg., ca 1275; MS1 (transcr. inédite R. L. H. Lops).

43: En fig. 9.19.d, $p = 0$ est un défaut d'arondi de notre part; lire $p < 0,005$.

pronoms toniques soient de moins en moins nombreux, vis-à-vis des pronoms atones. Étant donné que les premières formes élidées précèdent la séparation entre série tonique et série atone (élision métrique dans *Roland* puis graphique à partir du *Lapidaire* de Philippe de Thaon, 1^{er} t. XII^e s.; cf. sec. 9.1.1.1), il est au demeurant impossible de séparer, dans l'absolu, les deux séries. MARCHELLO-NIZIA (2020c, GGHF, p. 679) semble pencher en faveur d'une combinaison des deux hypothèses : d'une part l'élision devient de plus en plus systématique dans des positions traditionnellement faibles (devant le verbe, hors emploi emphatique), d'autre part elle se développe dans de nouveaux contextes, alors qu'elle était à l'origine restreinte à la précession d'un verbe ou d'un autre pronom. Mais, parallèlement, l'ensemble des formes du pronom personnel sujet de 1sg se réduit à *je* (élide ou non), avec la disparition progressive des formes à accentuation ambiguë (*jou, jeo*, etc.) et des formes strictement toniques (*gie*; GGHF *id.*, p. 680). Le changement serait donc à la fois phonologique (croissance de l'élision) et morphosyntaxique (croissance de l'emploi de formes faibles au détriment de formes fortes).

9.2.6.2 Distribution diatopique

Du fait de l'absence patente de formes défectueuses devant initiale consonantique, nous ne commentons pas la distribution diatopique des données préconsonantiques. Devant initiale vocalique, l'élision présente une distribution diatopique curieuse, avec des taux supérieurs à la moyenne principalement concentrés à l'Ouest et sur un arc de cercle allant de l'Oise à la Meuse. Ainsi, en fig. 9.20 page suivante, on observe à l'Est des taux de 100% dans la région Indre, Cher et en Indre-et-Loire, un taux de 80% dans l'Orléanais et un taux de 76% en Normandie. Si les taux de ces trois premières régions ne sont pas consolidés, les régions ne disposant que d'un texte — or il existe de nombreux textes n'ayant que des formes élidées — le taux de la Normandie est en revanche appuyé sur 22 textes. Le taux de formes défectueuses atteint également 75% dans la Meuse (3 textes), 70% dans l'Oise (6 textes), ces deux régions étant reliées par deux régions dont le taux, inférieur à 70%, reste supérieur à la moyenne (Aisne : 63%, 11 textes; Ardennes, 56%, 5 textes). L'Angleterre ainsi que quelques régions isolées sur le continent témoignent de taux bien inférieurs à la moyenne, avec 15% en Angleterre (38 textes), 14% dans le Nord (2 textes), 17% en région Moselle, Meurthe-et-Moselle (3 textes), 23% en Charente-Maritime (1 texte) et 29% dans l'Yonne (5 textes).

Étant donné le tracé de la droite de régression en fig. 9.19.b,d, il est probable que tout ou partie de cette distribution diatopique ne soit due qu'à la distribution temporelle des textes parmi les régions. La section suivante invalide partiellement cette hypothèse.

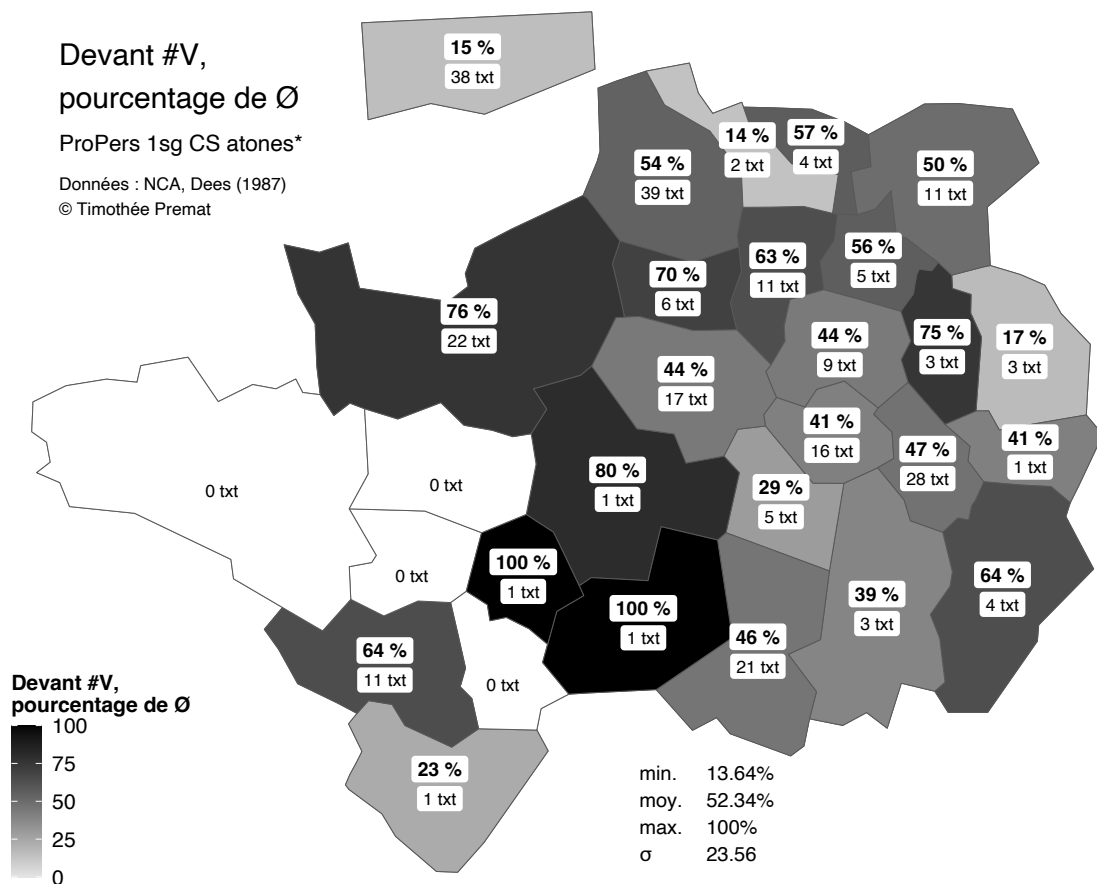


FIGURE 9.20 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, pronoms personnels sujets de première personne du singulier suivis d'une initiale vocalique.

*Atones : formes atones, formes indéterminées ou homographes.

9.2.6.3 Intersection de la diatopie et de la diachronie

Contrairement à ce que la croissance du taux de défaut devant initiale vocalique peut laisser supposer, la distribution diatopique observée en synchronie n'est pas uniquement due à la date de composition des textes. Les cartes diachroniques sont présentées en fig. 9.21 et 9.22, en pages suivantes. La moyenne des taux régionaux, telle qu'indiquée sous chaque carte, suit bien une croissance continue analogue à celle figurée par les régressions en fig. 9.19.b,d, passant de 43% au début du corpus à 59% à la fin de celui-ci, ce qui plaide en faveur de l'origine purement diachronique de la diatopie synchronique observée. Néanmoins, certaines régions se maintiennent au-dessus du taux moyen à toutes les époques : c'est le cas de la Normandie, qui se maintient toujours 17% ou plus au-dessus du taux régional moyen. C'est également le cas de l'Oise, tandis que l'Aisne dispose d'un taux bien supérieur au taux régional moyen à toutes les périodes sauf dans la première moitié du ^{xiii}e s. (fig. 9.22.a). En parallèle négatif de la situation de la Normandie et de l'Oise, l'Angleterre se maintient toujours bien en-deçà du taux régional moyen, tandis que

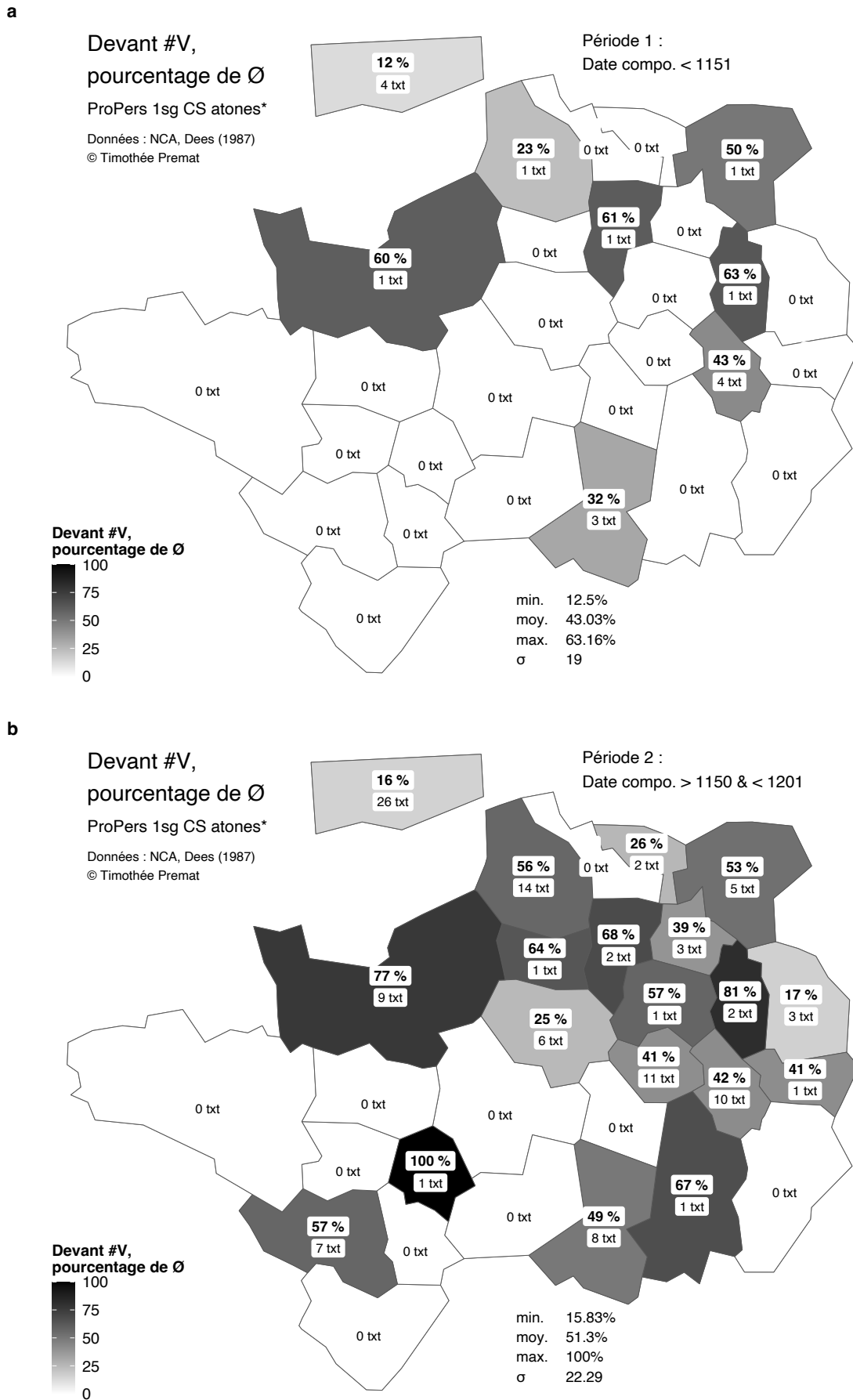


FIGURE 9.21 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1050) et 2 (b : 1151-1200), pronoms personnels sujets de première personne du singulier devant initiale vocalique.

*Atones : formes atones, formes indéterminées ou homographes.

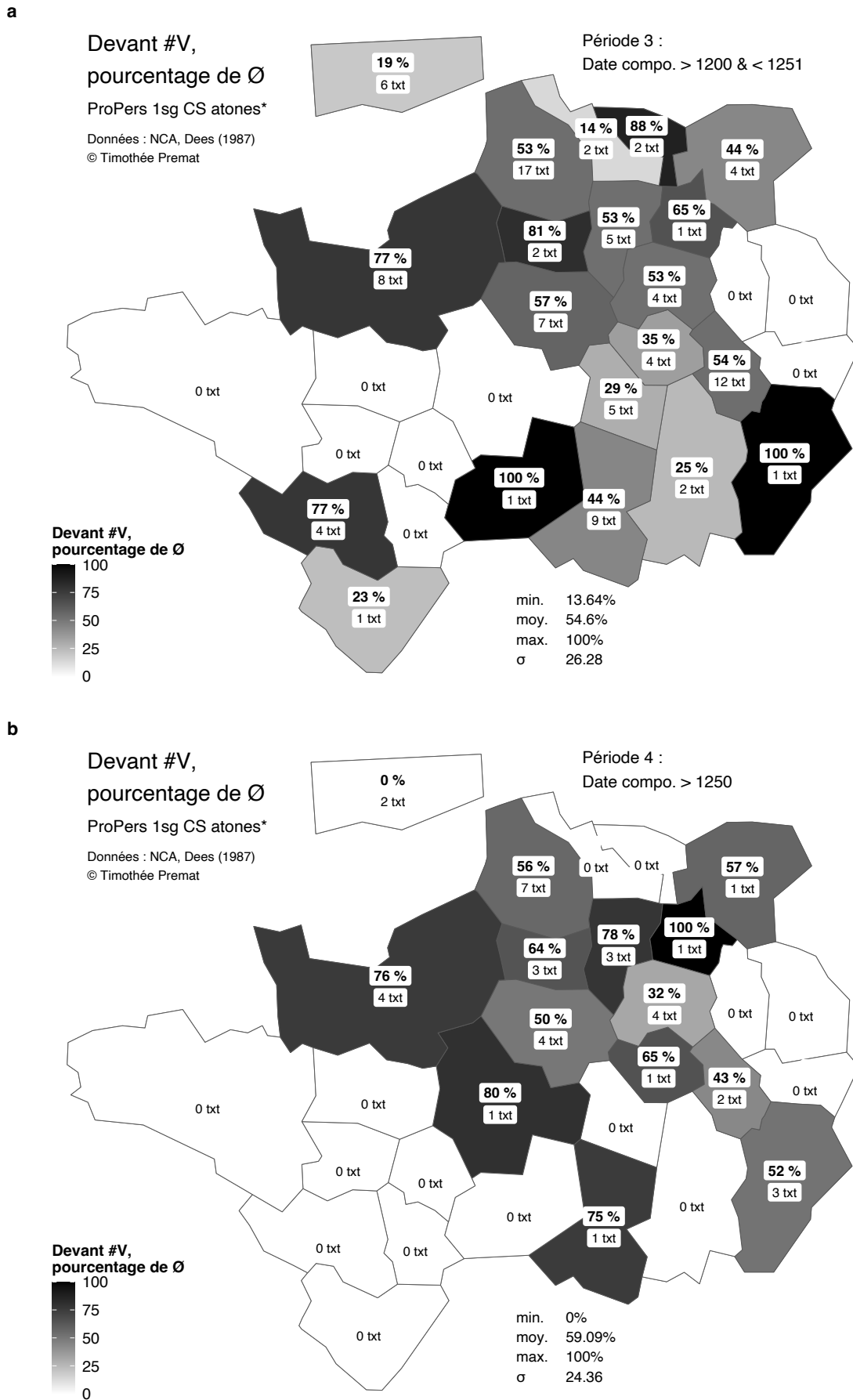


FIGURE 9.22 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), pronoms personnels sujets de première personne du singulier devant initiale vocalique.

*Atones : formes atones, formes indéterminées ou homographes.

des régions comme l'Yonne ou le Nord n'apparaissent qu'au ^{xiii} s., et ne peuvent donc avoir leur faible taux de défaut à la précocité de leurs textes. La seule diachronie ne saurait donc expliquer la distribution diatopique, le domaine étant peu homogène à toutes les époques (σ compris entre 19 et 26); l'Angleterre se caractérise par un taux extrêmement bas à toutes époques, tandis que la Normandie se caractérise, au contraire, par des taux élevés à toutes les époques.

9.2.6.4 En somme.

Le pronom personnel sujet de 1sg ne connaît pas de cas d'enclise généralisée. Devant initiale vocalique, le taux de défaut, qui porte sur l'ensemble des formes pouvant être toniques ou atones, est en croissance continue, mais avec une large distribution des données (de 100% de *j* à 100% de *je*, *jou*, *jeo* etc.). Cette croissance continue du taux de défaut semble résulter de deux forces conjointes : d'une part l'extension de l'élision elle-même, qui se libère des contextes syntaxiques précis dans lesquels elle est née, et d'autre part la réduction de la part des formes toniques. La distribution diatopique de ces formes indique que l'Angleterre et quelques régions continentales isolées pratiquent très peu l'élision, tandis que l'élision domine largement sur un grand arc de cercle partant de la région Indre, Cher, passant par la Normandie et l'Oise, et allant jusqu'à la Meuse. Si certaines de ces régions n'ont des taux de *j* anormalement bas ou élevés que par la distribution diachronique de leurs textes (précoces pour les taux bas, tardifs pour les taux élevés), ce n'est pas le cas pour toutes les régions, et en particulier pour l'Angleterre et la Normandie, qui demeurent respectivement bien en-deçà et bien au-dessus des taux moyens de défaut à toutes les époques.

9.2.7 PROPERSCS2SG : *tu, te* ~ *t*

Dans le tableau du paradigme des pronoms personnels que propose MARCHELLO-NIZIA (2020c, GGHF, p. 676-677) et dont nous reproduisons le contenu en table 9.1 *supra*, l'autrice note que le pronom sujet de deuxième personne du singulier peut, en position atone, être sujet à un défaut de voyelle. Dans nos données, ces cas sont relativement rares, et si nous n'avons pas été ici en mesure d'isoler les formes toniques des formes atones (*tu* appartient aux deux séries), ce simple critère ne suffit probablement pas à expliquer une telle rareté des défauts de *tu*, *te*. En effet, alors que le pronom sujet de 1sg était concerné par la même impossibilité d'isoler formes toniques et atones, son taux moyen de défaut était de 48%; pour le pronom sujet de 2sg, ce taux n'est que de 1% (cf. sec. suivante). Il semble donc nécessaire de mobiliser une autre motivation que l'absence de filtre entre formes toniques et atones. La situation, sur ce point, est analogue à ce qui est rencontré en fr.mod., où l'élision du pronom sujet de 2sg n'est en principe pas transcrite à l'écrit et est stylistiquement marquée à l'oral (*tu aimes* ~ *t'aimes*).

9.2.7.1 Distribution contextuelle et diachronique

Les graphiques de séries temporelles en 9.23.a-d, page suivante, représentent le taux de défaut du pronom personnel sujet de deuxième personne du singulier, devant initiale consonantique (a,c) et devant initiale vocalique (b,d), avec pondération des taux par le nombre d'occurrences des textes (c,d) ou sans pondération (a,b).

t __#C? Comparalem n'a détecté aucune occurrence *t* devant initiale consonantique. Ce résultat peut s'expliquer par l'absence d'enclise impliquant le pronom sujet de 2sg, par le fait que de telles enclises ne seraient pas écrites en deux mots dans le corpus, et/ou par le fait que de telles occurrences *t* auraient été annotées comme des cas régimes dans le corpus. Le fait que Comparalem ait bien détecté des occurrences *t* sujet devant initiale vocalique et qu'il n'ait détecté que marginalement des occurrences *t* régime devant initiale consonantique (sec. 9.2.4.1), nous incite à penser qu'il n'y a pas de problème général d'annotation en cas des occurrences *t*, et que le dernier motif peut donc être évacué. De fait, les quelques occurrences que nous avons pu détecter pour *t* régime devant initiale consonantique représentaient des cas marginaux d'enclise ou des occurrences suspectes, et, parmi les grammaires que nous avons consultées, seul WALKER (1981, p. 51) mentionne une enclise du pronom de 2sg, *jat*, sans préciser s'il s'agit du pronom régime ou sujet. Ainsi, le pronom de 2sg semble rétif à l'enclise : non seulement celle-ci est extrêmement rare au CR, mais elle semble inexistante au CS.

t __#V. Devant initiale vocalique, Comparalem a détecté 13 textes où le pronom sujet de 2sg se réduit parfois à *t*. Ces 13 textes représentent 18 occurrences, soit à peine plus d'une occurrence *t* par texte. Néanmoins, la majorité de ces 13 occurrences sont au mieux suspectes, au pire clairement fausses au niveau de l'annotation, et *t* y est toujours régime et non sujet, comme en (9.54 - 9.55). Dans d'autres cas, l'occurrence semble bien être le pronom sujet, comme en (9.56), mais ces cas sont minoritaires. Aussi, non seulement les occurrences de *t* sujet détectées par Comparalem sont rares (13 sur 1486 devant initiale vocalique), mais en plus la majorité d'entre elles découle d'une erreur d'annotation. Face à cette extrême marginalité, nous concluons que les cas d'élision du pronom sujet de 2sg sont des exceptions, et que la règle est que ce pronom résiste à l'élision. C'est, du moins, ce dont témoigne notre corpus et son annotation.

(9.54) *voiz tu ami ce que ie t auoie promis*
« Vois-tu, ami, ce que je t'avais promis ? »
(fablesC, p. 26, trad. pers.)

(9.55) *ci que t en semble*
« ce qui te semble à ce propos »
(fablesY, p. 21, trad. pers.)

(9.56) *li lecon ensegnet [...] ou t entendes mont profetes*
« les leçons enseignées [...] où tu acquières la compréhension de nombreux prophètes »
(beat i, p. 27, trad. pers.)

fablesC : MARIE DE FRANCE (1180*). *Fables*. Ms. C; Dees : nil; compo. NO, ca 1180; ms. lorr., ca 1275; MS (transcription inédite).

fablesY : MARIE DE FRANCE (1180*). *Fables*. Ms. Y; Dees [Angleterre]; compo. NO, ca 1180; ms. norm. et a.n., ca 1225; MS (transcription inédite).

beat i : ANON. (1200*). *Dialogus anime conquerentis et rationis consolantis*. Dees : Vosges; compo. lorr., ca 1200; ms. lorr., ca 1200; MS2 (BONNARDOT 1876); DEAF : DialAmeB.

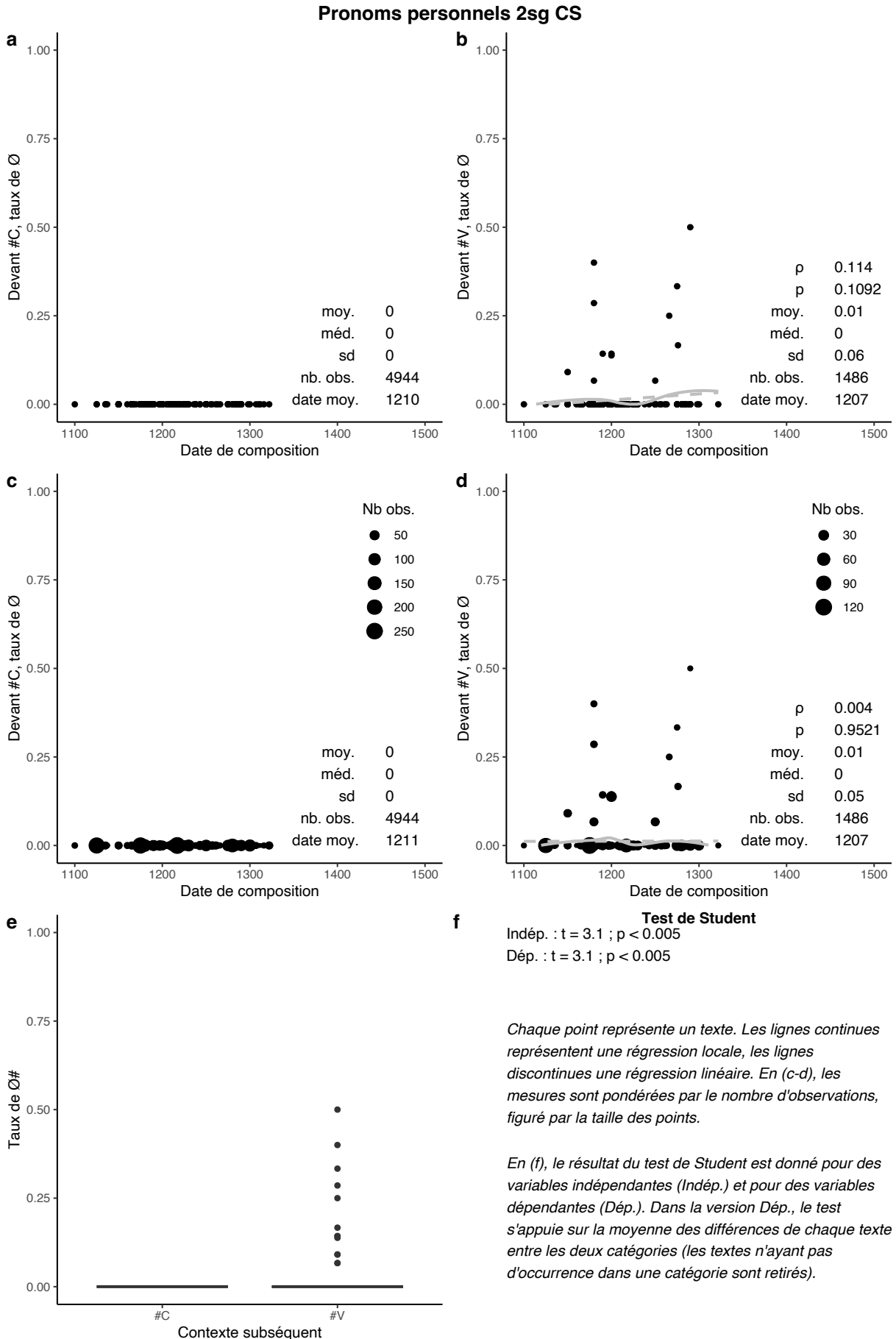


FIGURE 9.23 – Distribution contextuelle des variantes de pronoms personnels de deuxième personne du singulier au cas sujet en défaut de <e>.

9.2.7.2 En somme

Notre corpus n'atteste aucune occurrence *t* sujet devant initiale consonantique et les quelques attestations de *t* sujet devant initiale vocalique qui ne sont pas à rejeter sont extrêmement marginales. Aussi, nous concluons que le corpus et son annotation ne permettent pas d'attester d'une véritable élision ou enclise du pronom personnel sujet de deuxième personne du singulier.

9.2.8 PROPERSFÉMCS3sg : *ele* ~ *el*

Les graphiques de séries temporelles en fig. 9.24.a-d, page suivante, représentent le taux d'occurrences *el* du pronom personnel sujet de troisième personne du singulier au féminin.

9.2.8.1 Distribution contextuelle et diachronique

Le pronom sujet de FÉM 3sg ne reproduit ni la distribution des pronoms pour lesquels le contexte prévocalique autorise le défaut (CR : 1sg *m*, 2sg *t*, 3sg *l*; CS : 1sg *j*) ni celle du pronom sujet de 2sg, pour lequel notre corpus ne valide pas l'existence de formes défectueuses. En effet, les formes *el* sont attestées dans les deux contextes, et sont même deux fois plus nombreuses devant initiale consonantique que devant initiale vocalique, proportionnellement aux formes *ele*.

el __#C. Devant initiale vocalique, le taux moyen de formes *el* est de 14%. La tendance générale représentée par la droite de régression en fig. 9.24.a,c est légèrement à la baisse, mais la régression locale indique que cette tendance n'est guère fiable. Dans plus de la moitié des textes, seule la forme non défectueuse est présente (méd. = 0), et, comme indiqué par la boîte à moustaches en fig. 9.24.e, les textes qui comportent plus de 50% de formes *el* sont marginaux.⁴⁴ Ce graphique montre néanmoins qu'une proportion non négligeable de textes comporte ces formes *el*, dans la mesure où le troisième quartile s'étire jusqu'à ~20%. Les formes *el* sont donc minoritaires dans une large majorité des textes, mais sont tout de même présentes dans une minorité assez large des textes.

el __#V. Devant initiale vocalique, le taux moyen de formes *el* chute à 6%. Comme représenté sur la boîte à moustache, tous les textes comportant des formes *el* sont considérés comme dotés de valeurs extrêmes, dans la mesure où les trois quarts des textes ne disposent que de la forme *ele* devant initiale vocalique. Comme indiqué par les régressions locales en fig. 9.24.b,d, les textes comportant des formes *el* sont concentrés au centre du corpus, du milieu du XII^e s. à la seconde moitié du XIII^e s.

L'effet du contexte est faible ($t = -4,2$) mais statistiquement significatif ($p < 0,005$). Contrairement à ce que nous avons noté pour les pronoms régimes de 1-3sg et pour le pronom sujet de 1sg, cet effet est négatif. La situation du pronom sujet *ele* est donc proche, sur ce critère, de celle que nous avons notée *supra* pour les adverbes (Student Dép. $t = -2,5$; Student Indép. $t = -4,5$; cf. fig. 7.8, p. 269). Il convient donc de considérer, comme

44: Leur taux les classe comme des *outliers*, puisqu'ils sont séparés de la limite du troisième quartile par plus de 1,5 fois l'écart inter-quartile.

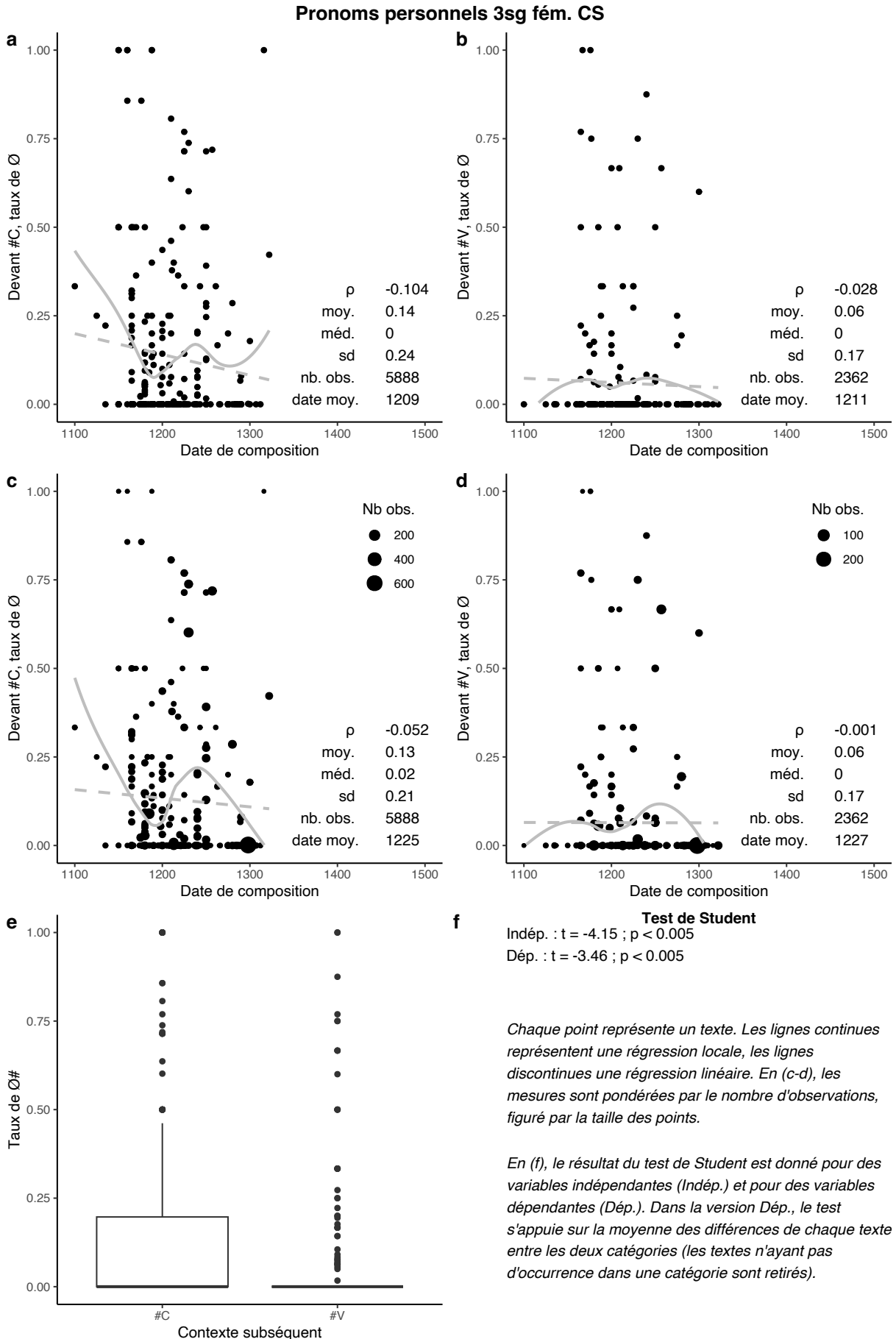


FIGURE 9.24 – Distribution contextuelle des variantes de pronoms personnels de féminin troisième personne du singulier au cas sujet en défaut de <e>.

pour les adverbes, que la variation à l'œuvre ici n'est pas conditionnée par le contexte, mais paraît libre. C'est d'ailleurs ce que semblent suggérer les grammaires qui, comme MARCHELLO-NIZIA (2020c, GGHF, p. 676-677), mentionnent la forme *el* comme une simple variante de *ele*. Néanmoins, à la différence des adverbes pour lesquels nous avons noté cette variation, la forme défectueuse est ici minoritaire (14% __#C, 6% __#V) alors qu'elle était majoritaire pour les adverbes (74% __#C, 68% __#V).

9.2.8.2 Distribution diatopique

Que ce soit devant initiale consonantique ou vocalique, les formes *el* semblent se concentrer au Sud-Ouest, en-deçà d'une ligne reliant la Normandie à la région Indre, Cher (devant #C) ou à la Bourgogne (devant #V). Comme indiqué sur les cartes synchroniques en fig. 9.25, en page suivante, la situation est moins homogène devant initiale consonantique ($\sigma = 21$) que devant initiale vocalique ($\sigma = 14$). Parallèlement, alors que devant initiale vocalique, aucune région au Nord-Ouest de l'axe Normandie-Bourgogne ne dispose de taux supérieurs ou égaux au taux régional moyen, ce n'est pas le cas devant initiale consonantique, avec notamment 21% en région parisienne (17 textes), 17% dans l'Aisne (11 textes) et 11% en Haute-Marne (28 textes).

Par ailleurs, si *el* semble bien être une forme liée au Sud-Ouest du domaine, elle ne domine pas même au sein de cette région : les taux très élevés constatés dans l'Orléanais (72% __#C, 67% __#V) et en région Indre, Cher (77% __#C) ne valent que pour un texte et ne sont donc pas consolidés. Non seulement toutes les autres régions en-deçà de l'axe décrit disposent de taux inférieurs à 50%, mais certaines disposent aussi de taux inférieurs à la moyenne, avec 0% dans les deux contextes en Charente-Maritime et en Indre-et-Loire, certes sur un seul texte par région.

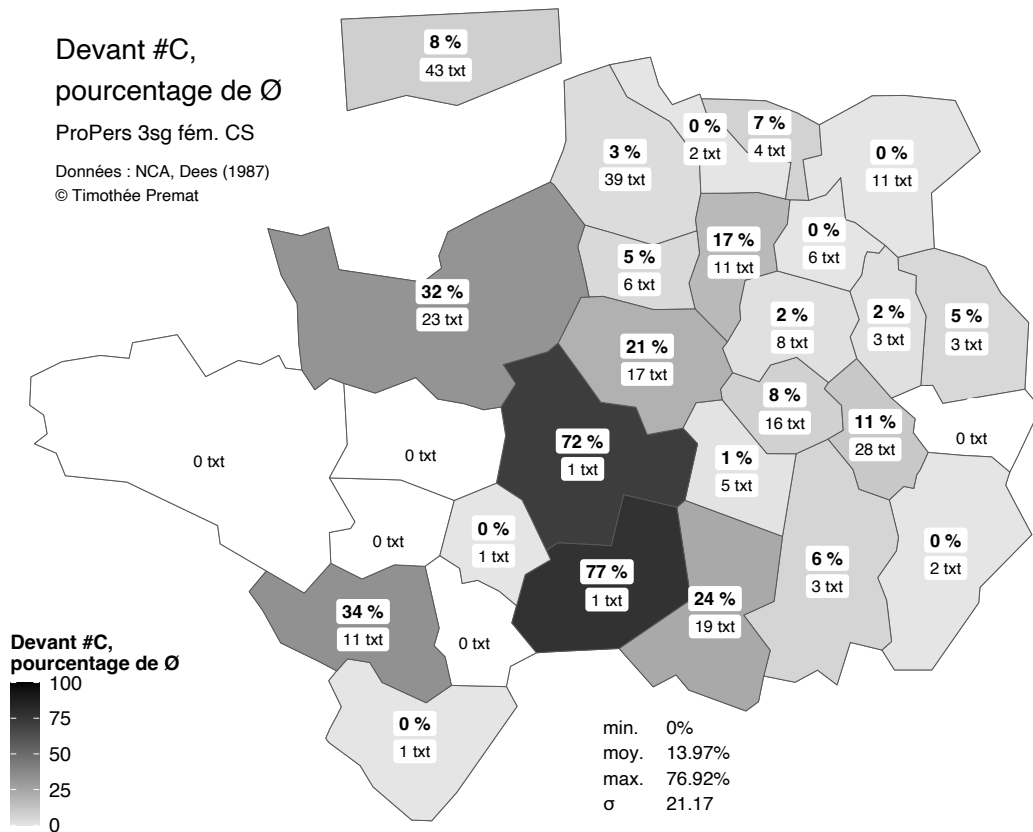
9.2.8.3 Intersection de la diatopie et de la diachronie

Nous ne produisons pas, ici, de commentaire détaillé de toutes les cartes diachroniques, qui sont rejetées en annexes (fig. E.14 - E.17, p. 623 - 626). Nous nous contentons de remarquer que cette distribution orientée vers le Sud-Ouest des formes *el* est relativement stable dans l'espace. Cette stabilité n'est que relative, dans la mesure où les moyennes étant basses, des régions peuvent facilement les dépasser lorsqu'elles disposent, à une période donnée, d'un seul texte. Néanmoins, la forme *el* s'affirme bien comme une variante minoritaire et non contextuelle de *ele*, plus fortement présente au Sud-Ouest du domaine qu'ailleurs.

9.2.8.4 En somme

La variation *ele* ~ *el* ne semble pas liée au contexte, et ne partage aucune caractéristique avec les défauts de voyelle finale atone observés pour les autres pronoms personnels. Plutôt que de se comporter comme les autres pronoms, *el(e)* semble se comporter comme les adverbes à variation <e>/Ø, c'est-à-dire être doté d'une variation libre. Sa distribution diatopique est néanmoins marquée, avec des formes *el* bien plus nombreuses

a



b

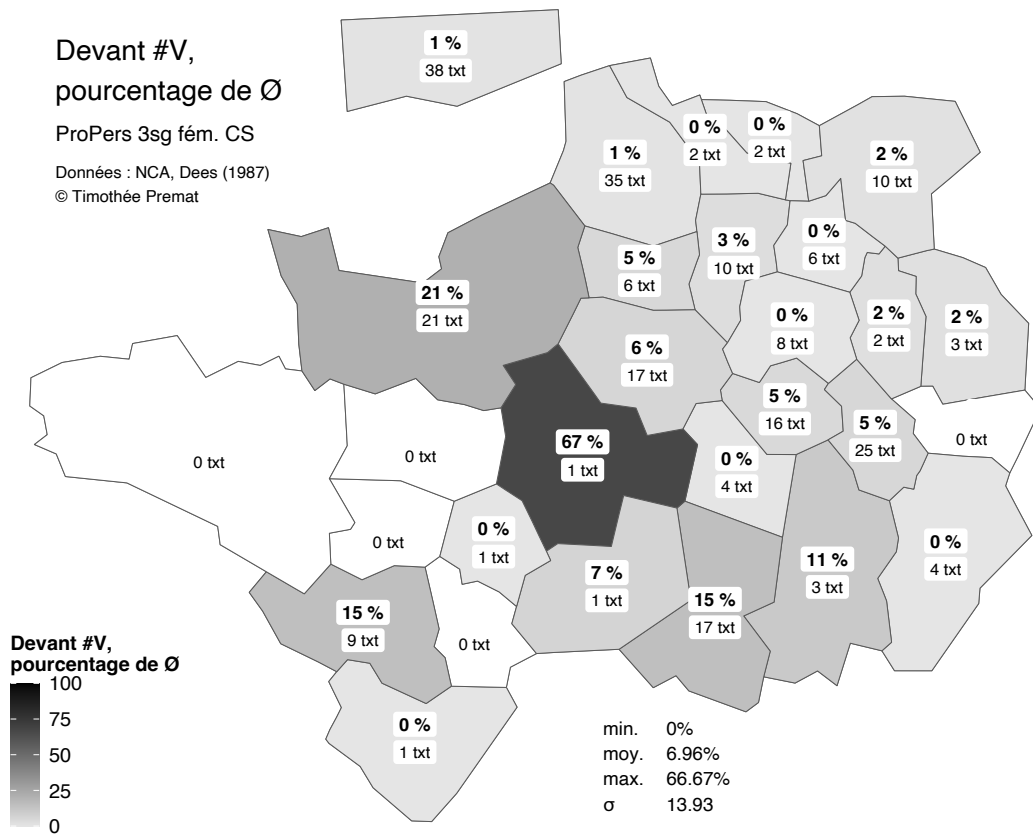


FIGURE 9.25 – Distribution diatopique des alternances <e>/Ø, pronoms personnels sujets féminins de troisième personne du singulier devant initiale consonantique (a) et devant initiale vocalique (b).

au Sud-Ouest du domaine qu'ailleurs, mais toujours minoritaires si l'on s'en tient aux résultats consolidés par un nombre de textes satisfaisant, et cette distribution est relativement stable dans le temps.

9.3 Des observations aux généralisations

Dans cette section, nous tentons de regrouper les observations faites *supra* pour chaque pronom étudié, pour obtenir des généralisations pertinentes. Nous commençons par constater qu'aucun critère phonologique, morphologique ou morphosyntaxique ne peut, à lui seul, prédire les défauts de <e>, ce qui nous amène à admettre une sélectivité lexicale de la cible des processus (sec. 9.3.1). Nous reprenons ensuite l'effet du contexte, qui prend ici la forme d'une distinction entre élision, enclise et libre variation (sec. 9.3.2), avant d'aborder la diachronie (sec. 9.3.3) et les quelques éléments de distribution diatopique (sec. 9.3.4). Nous faisons ici l'impasse sur la place de la variation au sein de la catégorie, étant données la spécificité de la catégorie en question et la répartition très inégale de la variation entre les cas et personnes.

9.3.1 Lemmes atteints par l'enclise et l'élision

Tous les pronoms personnels ne sont pas atteints par l'enclise et l'élision. L'enclise, c'est-à-dire la réduction du pronom à sa seule consonne sous l'effet de sa fusion avec le mot précédent, ne concerne guère que le pronom régime masculin de 3sg *le*; pour les autres pronoms, les occurrences consonantiques devant initiale consonantique sont soit absentes soit extrêmement rares. L'élision, quant à elle, concerne tous les pronoms régimes singuliers — nous n'avons pas pu l'attester au féminin de la troisième personne à cause d'un défaut récurrent d'annotation du corpus, mais il y a lieu de supposer qu'elle y est tout aussi active qu'au masculin. L'élision concerne aussi le pronom personnel sujet de 1sg *jo*, *je*, mais comme nous n'avons pas pu séparer les formes toniques des formes atones, nous n'avons pas pu établir si elle était systématique ou non; MARCHELLO-NIZIA (2015) indique qu'elle reste longtemps optionnelle, si tant est qu'on puisse savoir que le pronom en question est faible. En revanche, dans nos données, l'élision ne concerne pas le pronom sujet de 2sg *tu*, *te*. Enfin, ni l'élision ni l'enclise ne semblent concerner le pronom sujet de 3sg féminin *ele*, dont la variation ressemble à celle des adverbes et est surtout diatopiquement marquée au Sud-Ouest du domaine.

Dans les données de MARCHELLO-NIZIA (2020c, GGHF, p. 676-677), l'élision ne concerne jamais les formes du cas régime en construction indirecte (*li*, *en*, *i*, *lui*). Dans une exécution plus ample de Comparalem que celle que nous avons présentée ici, et analogue à celle déployée pour les autres parties du discours, le programme n'a pas non plus détecté de défauts sur ces formes. L'élision ne concerne pas non plus les pronoms pluriels, d'une part parce que les 1pl et 2pl forment une série isolée au sein du paradigme pronominal, d'autre part parce que les 3pl MASC/FÉM sont dotés d'une consonne finale qui interdit l'élision. Enfin, selon les grammaires, le pronom réfléchi de 3sg-pl atone *se* est sujet à élision, mais nous n'avons pas vérifié cet élément.

Il y a donc un effet de sélectivité lexicale : on ne peut pas, sur la base de critères purement phonologiques, décider quelles seront les formes atteintes : le pronom régime de 3sg FÉM *la* devrait être autant sujet à l'élision que le masculin *le*, et les pronoms atones sujets ne se terminent pas nécessairement par un <e> pour schwa.⁴⁵ La morphologie elle-même ne semble guère cohérente, dans la mesure où elle admettrait l'élision pour le pronom sujet de 1sg mais pas pour celui de 2sg. Enfin, syntaxiquement, la structure *PROPERSCS + verbe* (qui est de très loin la plus courante) est la même quel que soit le pronom, et ne peut donc traiter différemment les 1sg CS et le 2sg CS. Ces éléments sont synthétisés en (9.57) et (9.58).

45: Et ce <e> ne désigne pas nécessairement un schwa, cf. *naje, najo* (AND « not I! »).

(9.57) **PROPERs : sélectivité lexicale.**

Il est impossible, sur la base de critères phonologiques, de prédire quels pronoms sont atteints par l'enclise et l'élision.

(9.58) **PROPERs : sélectivité morpho-syntaxique.**

Ni la morphologie ni la syntaxe ne sont en mesure de prédire quels pronoms sont atteints par l'élision.

Il est en revanche remarquable que tous les pronoms atteints par l'élision dès le début du corpus sont des pronoms régimes, et que le seul pronom sujet concerné ne développe que plus tardivement la distinction entre série atone et tonique et, parallèlement, sa capacité à l'élision. Il semble donc qu'il y ait une transmission de l'élision, depuis les pronoms régimes singuliers, vers le pronom sujet de 1sg. Si l'on admet cette influence des pronoms régimes, il suffit de postuler que la contamination a affecté le pronom sujet de 1sg mais n'est pas parvenue à atteindre réellement le pronom sujet de 2sg. L'absence de cohérence interne, et donc de prédictibilité de l'élision, serait alors un simple effet de la dimension stochastique de la diffusion d'un traitement phonologique. Quant au pronom sujet féminin de troisième personne, il semble suivre sa propre logique et, en l'occurrence, peut être distingué sur un fondement phonologique : si l'on admet une représentation sous-jacente et/ou diasystémique bisyllabique /ɛlə/, il s'agit du seul pronom à ne pas être monosyllabique. Ces éléments sont repris en (9.59) et (9.60).

(9.59) **PROPERs : diffusion lexicale.**

L'élision est présente d'emblée pour les pronoms régimes, et ne se développe que progressivement pour le pronom sujet de 1sg.

(9.60) **PROPers : restrictivité phonologique.**

Si la seule phonologie ne peut prédire quels seront les pronoms personnels sujets à l'élision, elle semble en revanche fournir un critère d'exclusion du pronom sujet féminin de 3sg, qui est le seul à être bisyllabique (*elle*). Ce pronom se comporte comme les adverbes.

Enfin, pour ce qui est de l'enclise, que l'on peut résumer avec LA CHAUSSÉE (1989 [1974], p. 120-121) comme une perte de voyelle finale du second élément, elle ne semble s'appliquer réellement qu'au pronom régime de 3sg. Cet élément, bien présent dans nos données (3sg CR : 18,5% __#C vs autres pronoms < 1%; sauf *elle*), ne fait que confirmer les listes de formes données par les grammaires, qui soit n'indiquent pas d'enclise avec d'autres seconds éléments que le pronom régime de 3sg,⁴⁶ soit les indiquent mais en précisant que celles-ci sont rares et archaïques.⁴⁷ La sélectivité du seul pronom régime de 3sg, ici, est peut-être liée à la fréquence des constructions, dans la mesure où le pronom régime de 3sg est le plus courant des pronoms régimes (cf. table 9.2 *supra*); la différence n'est cependant pas flagrante (devant #C, PROPersCR3sg : ~13 500 occ.⁴⁸ vs PROPersCR1sg : 9746 occ.). Néanmoins, dans la mesure où les supports d'enclise sont de plus en plus restreints et où les structures enclitiques perdent peu à peu leur transparence et deviennent lexicalisées lorsqu'elles sont devenues opaques (FOUCHÉ 1966, p. 590; POPE 1966 [1934], p. 323), un processus de grammaticalisation semble à l'œuvre. La sélectivité de celui-ci, si elle peut nécessiter une base fréquentielle (dont elle dispose ici), semble par nature aléatoire. La formulation en (9.61) reprend ces éléments. Cette remarque remplace celle que nous tenons sur les collocations pour les autres parties du discours.

(9.61) **PROPers (enclise) : grammaticalisation.**

L'enclise ne concerne réellement que le pronom régime de 3sg, probablement suite à un processus de grammaticalisation. La fréquence d'emploi des pronoms régimes de 3sg est compatible avec une telle grammaticalisation.

46: CLÉDAT 1917, p. 1917; FOUCHÉ 1969 [1958], p. 429, 440, 1966, p. 590, 653, 859; LA CHAUSSÉE 1989 [1974], p. 120-121; LUQUIENS 1919 [1909], p. 88; PRICE 1985 [1971], p. 115-116.

47: NYROP 1899, p. 240; POPE 1966 [1934], p. 323; WALKER 1981, p. 51, 64.

48: Rappelons que quelques milliers de formes féminines sont annotées masculines.

9.3.2 Le rôle du contexte

Le contexte joue un rôle majeur dans les défauts de tous les pronoms personnels singuliers au cas régime et du pronom personnel sujet de 1sg. En effet, l'élision, devant initiale vocalique, est presque systématique pour les pronoms régimes (entre 96,2 et 98,6%), et très présente pour le pronom sujet de 1sg (45,9%). Pour les pronoms qui ne connaissent pas l'enclise, le contexte prévocalique joue le rôle de condition nécessaire et, pour les pronoms régimes, cette condition semble aussi suffisante, provoquant une élision presque systématique. Pour le pronom régime de 3sg, fortement sujet à l'enclise (18,5% __#C), le contexte prévocalique n'est pas une condition nécessaire en soi, puisque le pronom peut aussi perdre sa voyelle sous l'effet d'un processus de fusion avec le mot précédent. Ainsi, les deux contextes subséquents possibles divisent les taux positifs de défaut entre deux processus phonologiques distincts, l'élision et l'enclise, et ces deux processus n'ont pas les mêmes cibles. En revanche, le contexte n'a évidemment un effet que très limité voire nul sur les pronoms personnels qui ne connaissent ni enclise ni élision (2sg CS)

ou qui suivent une autre logique de variation <e>/Ø (3sg FÉMCS). Les configurations où le contexte joue un rôle, et la façon dont il joue ce rôle, sont résumés en (9.62) - (9.64); les mesures indiquées sont celles des jeux de données non pondérés par le nombre d'occurrences des textes et des tests de Student à variables indépendantes.⁴⁹

49: Cf. fig. 9.11, 9.5, 9.1, 9.19.

(9.62) **ProPERS : contexte, 1-2sg CR : *me, te***

Le contexte prévocalique agit comme condition nécessaire et presque suffisante au déclenchement du défaut des pronoms régimes de 1 (*m(e)*) et 2sg (*t(e)*).

(1sg CR : $t = 155$; $p < 0,005$; moy. $_\#V = 97\%$)

(2sg CR : $t = 92$; $p < 0,005$; moy. $_\#V = 96\%$)

(9.63) **ProPERS : contexte, 3sg CR (MASC.) : *le***

Le contexte prévocalique agit comme condition suffisante mais non nécessaire au déclenchement du défaut des pronoms régimes de 3sg (*l(e)*), dans la mesure où ce pronom connaît aussi l'enclise devant initiale consonantique.

($t = 58$; $p < 0,005$; moy. $_\#V = 99\%$; moy. $_\#C = 21\%$)

L'élision est stable dans le temps, l'enclise décroît; cf. (9.65).

(9.64) **ProPERS : contexte, 1sg CS : *je, jo***

Le contexte prévocalique agit comme condition nécessaire mais non suffisante au déclenchement du défaut des pronoms sujet de 1sg (*j(e) etc.*), ce qui est dû à l'absence de discrimination des formes atones et à l'extension seulement progressive des contextes syntaxiques de l'élision.

($t = 22$; $p < 0,005$; moy. $_\#V = 48\%$)

9.3.3 Diachronie générale

La distribution diachronique des défauts des pronoms n'est pas unifiée : alors que le pronom sujet de 1sg ne connaît pas d'élision graphique dans *Roland* (mais connaît des élisions métriques), les défauts devant initiale vocalique des pronoms régimes sont stables dans le temps, et l'absence de défauts conséquents des pronoms sujets de 2sg l'est également. Pour le pronom régime de 3sg, Comparalem détecte une baisse progressive des défauts devant initiale consonantique (enclise), qui peut être en partie due au fait que le programme n'est sensible qu'aux enclises divisées en deux mots, alors que ces cas se font de plus en plus rares par suite de modifications phonétiques (type *a l* → *au*, *de l* → *dou*); néanmoins, la raréfaction de l'enclise est attestée par la littérature à ce sujet.⁵⁰ Quant aux défauts de <e> du pronom sujet féminin de 3sg, ils varient fortement dans le temps, mais étant donnée la dominante spatiale de leur distribution, cette variation diachronique est difficile à analyser. Ces éléments sont repris en (9.65).

50: LUQUIENS 1919 [1909], p. 83-84; NYROP 1899, p. 239-240; POPE 1966 [1934], p. 323.

(9.65) **PROPers : diachronie.**

L'élision et son évolution dépend du pronom impliqué (a-c), et l'enclise du pronom régime de 3sg est en régression (d). La variation du pronom sujet fém. 3sg est surtout diatopique (e).

- | | | |
|------------------|-----------|----------------------------|
| a) CR sg. | __#V | stable : 96-99% |
| b) CS 1sg. | __#V | croissance de 0 à ~60% |
| c) CS 2sg. | __#{C, V} | stable : 0% |
| d) CR 3sg (MASC) | __#C | décroissance de ~40% à ~0% |
| e) CS 3sg. FÉM | __#{C, V} | tendance à la décroissance |

9.3.4 Diatopie générale

La diatopie de l'élision des pronoms régimes (sauf *ele*) et des pronoms sujets de 2sg est évidemment neutre, dans la mesure où leurs taux sont proches de 100 ou de 0%. En revanche, pour les pronoms dont les taux de défaut ne sont pas proches de 100 ou de 0%, on peut observer des faits de distribution diatopique. D'une part, l'élision (devant #V) des pronoms de 1sg CS et les cas de défaut (élision comme enclise) du pronom sujet de 3sg FÉM se concentrent à l'Ouest du domaine, avec, pour l'élision du pronom de 1sg CR, une percée additionnelle sur un axe allant de la Normandie à la Meuse. D'autre part, l'enclise (devant #C) du pronom régime de MASC3sg se concentre au Nord-Est du domaine, du Nord aux Vosges. Nous avons donc affaire à deux distributions diamétralement opposées, ce qui ne plaide pas en faveur de l'unité du phénomène étudié. Enfin, l'anglo-normand joue cette fois un rôle mineur dans les variations observées : son taux ne se départ des taux continentaux significativement que pour l'élision (devant #V) du pronom sujet de 1sg ; or l'élision de ce pronom est en croissance et l'anglo-normand est précoce.

(9.66) **PROPers : diatopie.**

L'élision du pronom sujet de 1sg *j(e)* se concentre à l'Ouest et sur un axe de la Normandie à la Meuse (a), les défauts du pronom sujet de 3sg FÉM *el(e)*, devant #V comme devant #C, se concentrent également à l'Ouest, tandis que l'enclise du pronom régime de 3sg se concentre au Nord-Est (c). L'anglo-normand n'a pas de comportement général spécifique.

- | | | |
|------------------|-----------|-------------------------------------------------------|
| a) 1sg CS | __#V | Concentration à l'Ouest et de la Normandie à la Meuse |
| b) 3sg FÉM | __#{C, V} | Concentration à l'Ouest |
| c) 3sg CR (MASC) | __#C | Concentration au Nord-Est |

9.4 Conclusion

Dans ce chapitre, nous avons étudié la distribution des défauts de <e> des pronoms personnels dans le corpus du NCA, à l'exception du pronom personnel réfléchi. Après avoir, dans l'introduction (sec. 9.1), présenté le paradigme des pronoms personnels de l'a.fr. (sec. 9.1.1) et la situation spécifique du pronom sujet de première personne du singulier (sec. 9.1.1.1) et des pronoms régimes (sec. 9.1.1.2) au sein de ce paradigme, nous avons abordé le traitement des données, qui ne suit pas, ici, la chaîne de traitement habituelle (sec. 9.1.2). Nous avons

ensuite dressé une typologie ternaire des défauts de voyelle des pronoms personnels, pour aboutir à la conclusion que ce défaut ne concernait pas nécessairement uniquement des <e> pour schwas, et qu'il pouvait procéder de deux logiques distinctes : l'élision devant initiale vocalique et, indépendamment d'une telle initiale subséquente, l'enclise avec un terme support précédant immédiatement le pronom (sec. 9.2.2). Cela fait, nous avons étudié la situation de chaque pronom atteint par des défauts de voyelle finale atone, en commençant par les pronoms régimes de 3sg MASC (sec. 9.2.3), de 2sg (sec. 9.2.4) et de 1sg (sec. 9.2.5), avant d'étudier les défauts des pronoms sujets de 1sg (sec. 9.2.6), de 2sg (sec. 9.2.7) et enfin de 3sg. FÉM (sec. 9.2.8). Dans chaque cas, nous avons traité l'effet du contexte et, lorsque cela était pertinent, avons abordé les distributions diachroniques et diatopiques ainsi que leur intersection. Enfin, nous avons regroupé cet ensemble d'observations en 10 généralisations (sec. 9.3).

Ces données mettent en lumière le fait que l'élision ne concerne que les pronoms régimes singuliers (*me, te, le* — et a priori *la*) et le pronom sujet de 1sg (*je* et variantes). Contrairement à ce que semblait indiquer le tableau de la GGHF reproduit en table 9.1, l'élision du pronom sujet de 2sg *tu* est si rare qu'elle s'apparente plus au bruit de fond attendu pour des données d'une langue ancienne, non standardisée, et dont le corpus comporte des défauts d'annotation. L'enclise, quant à elle, ne concerne dans notre corpus que le pronom régime de 3sg *le*, et uniquement après un terme support qui appartienne à une liste réduite ; les autres cas sont trop isolés. Ces faits nous ont amené à déduire qu'un critère phonologique (9.57) ou morphosyntaxique (9.58) n'était pas à même, seul, de prédire quelle forme serait atteinte par l'élision et/ou l'enclise, et qu'il fallait plutôt analyser l'élision du pronom sujet de 1sg *j(e)* comme la diffusion, depuis une classe morphosyntaxique discrète (celle des pronoms régimes singuliers atones), d'un traitement phonologique. En revanche, si la phonologie ne peut pas prédire positivement les items qui seront affectés, elle peut prédire que certains items ne seront pas affectés, du moins pas de la même manière. Ainsi, *ele*, seul pronom personnel bisyllabique, se comporte comme un adverbe et non comme les autres pronoms personnels (9.60). L'enclise, quant à elle, ne peut faire l'objet d'une prédiction par classe, dans la mesure où elle ne concerne qu'un seul item — une fois l'époque des enclises archaïques révolues — et il faut donc se résoudre à l'analyser à l'aune d'une grammaticalisation, qu'on sait croissante par ailleurs alors que la phonétique vient modifier les formes et rendre les occurrences enclitiques opaques (9.61).

Le contexte joue donc, dans les occurrences défectueuses des pronoms régimes singuliers et du pronom sujet de 1sg, un rôle prépondérant (sec. 9.3.2). Le contexte prévocalique agit comme une condition non seulement nécessaire mais aussi suffisante pour les pronoms régimes de première et deuxième personne du singulier (*me, te*), dont l'élision est presque systématique et qui ne connaissent pas l'enclise (9.62). Pour le pronom régime de 3sg *le*, dans la mesure où celui-ci connaît l'enclise en plus de l'élision, le contexte prévocalique est une condition suffisante mais non nécessaire (9.63). Enfin, pour le pronom sujet de 1sg *je, jo* (etc.), le contexte prévocalique agit comme une condition nécessaire mais non suffisante (9.64), certes parce que nous n'avons pas pu isoler les formes atones, mais aussi parce que l'élision de ce pronom n'est pas systématique et ne gagne de l'ampleur que progressivement. Pour

ce qui est du pronom sujet de 3sg féminin *ele*, le contexte a un effet statistiquement significatif mais très faible et, comme pour les adverbes, nous avons conclu que cet effet était linguistiquement insignifiant. Sur le plan diachronique (sec. 9.3.3), les seules variations observées tiennent à la récession de l'enclise du pronom régime de 3sg et à la croissance de l'élision du pronom sujet de 1sg (9.65). Enfin, sur le plan diatopique, seuls trois pronoms présentent une distribution spatialement marquée : l'élision du pronom sujet de 1sg est plus forte à l'Ouest et sur un axe allant de la Normandie à la Meuse, les défauts de <e> du pronom sujet de 3sg féminin *ele* se concentrent plutôt à l'Ouest du domaine, tandis que les cas d'enclise du pronom régime de 3sg sont plutôt concentrés au Nord-Est du domaine (9.66).

La situation des voyelles finales des pronoms personnels est donc très variée en fonction du genre et de la personne et, dans le cas de l'enclise du pronom régime de 3sg et du pronom sujet de 1sg, des constructions syntaxiques. Deux dynamiques diachroniques contraires semblent s'opposer, avec d'une part l'extension de l'élision, qui cesse de ne concerner que les pronoms régimes et contamine le pronom sujet de 1sg progressivement, et d'autre part la lexicalisation croissante des cas d'enclise, qui va de pair avec une perte en productivité des constructions enclitiques. Ces deux dynamiques contraires ne sont pas concomitantes dans l'espace : l'enclise de *le* se maintient le plus là où l'élision de *je* n'est pas la plus forte et là où *ele* n'est pas le plus souvent réduit à *el*. Il est néanmoins indéniable qu'il ne s'agit pas d'une simple évolution vers plus de chute de schwas, mais de deux processus différents, qui évoluent vers deux horizons opposés.

Si l'on intègre à cette réflexion l'évolution morphosyntaxique de l'a.fr., qui voit notamment se diviser en série toniques et atones les formes des pronoms sujets — alors que les pronoms régimes sont, d'emblée, divisés ainsi — il convient d'introduire l'idée que tout ou partie de cette variation peut être due à des changements qui, tout en conditionnant la phonologie, ne sont pas eux-même phonologiques. En effet, si les formes toniques et atones se distinguent sur un fondement phonologique, leur motivation doit précéder la phonologie et, en l'occurrence, être syntaxique. Ainsi, à l'exception de la variation plus ou moins libre et diatopiquement marquée du pronom sujet *ele*, l'intégration du pronom sujet de 1sg au sein des items atteints par l'élision et, parallèlement, la perte de productivité de l'enclise, sont probablement à justifier hors du champ de la phonologie elle-même. L'impossibilité de décider, dans un seul module grammatical et sur la base de classes discrètes, des unités sujettes à l'élision ou à l'enclise fait signe vers une interaction forte entre les différents composants de la grammaire.

Conjonctions : *Que*, *-que* et *come*

10.1 Introduction

Ce chapitre traite des conjonctions, mais du fait de l'annotation du corpus, ses cibles sont fort réduites : ne sont concernées que les conjonctions complexes en *base* + *-que*, le mot *que* autonome, et la conjonction *come*. Toutes les autres conjonctions annotées comme telles dans le corpus ne disposent pas d'une variation <e>/∅ ou <V>/∅ détectable par Comparalem; ce chapitre ne porte donc que sur certaines locutions en *que*, sur *que* autonome et sur la conjonction *come*.

Après avoir fait un point sur les conjonctions de l'ancien français (10.1.1), nous abordons le statut morphosyntaxique de « que » seul (10.1.2) et la question de la distribution syntaxique entre les formes *com* et *come* (sec. 10.1.3), puis le traitement des données (10.1.4). Nous détaillons ensuite les données, en dédiant une section à chaque catégorie de lemmes, traitant d'abord des locutions conjonctives en *base* + *-que* (sec. 10.2.2), puis de « que » hors de ces locutions (sec. 10.2.3) avant de traiter de « come » (sec. 10.2.4), d'abord du seul point de vue de la distribution syntaxique (sec. 10.2.4.1) puis indépendamment de la nature du mot suivant (sec. 10.2.4.2). Du fait de l'hétérogénéité des données, les variables diachroniques et diatopiques ainsi que leur interaction sont traitées soit dans des sections dédiées (*base* + *-que*, « que » seul), soit dans des paragraphes au sein du développement (« come » selon le contexte syntaxique et dans tous les contextes syntaxiques). Comme dans les chapitres précédents, nous effectuons ensuite une montée en généralisation (sec. 10.3) avant de conclure (sec. 10.4).

10.1.1 Les conjonctions de l'ancien français

Le système conjonctif de l'a.fr. est composé de deux séries de conjonctions : les conjonctions simples et les locutions conjonctives. Au sein du système des conjonctions simples, la conjonction *que* est très présente, notamment parce qu'elle hérite à la fois des valeurs de QUOD et de QUIA, et que c'est elle qui sert à former la majorité des locutions conjonctives (*puis que*, *lors que*, *mal gré que*, etc.); les valeurs de *que* sont si larges que c'est le plus souvent le contexte qui définit leur précision (COMBETTES 2020, GGHF, p. 899). Quant aux locutions conjonctives, elles peuvent continuer des modèles latins en QUOD (PROPTER QUOD → *pour que*, AB EO QUE → *a ce que*, etc.) ou QUAM (PRIUS QUAM/ ANTEQUAM → *avant que*, POSTQUAM → *après que*), être formées par l'adjonction de *que* à un adverbe (*lors que*, *combien que*, *bien que*, etc.), à un participe (*moyennant que*,¹ *pendant que*, *vu que*, etc.) ou à un nom (*à condition que*, *à mesure que*; GGHF *id.*, p. 900-901). Dans bien des cas, la locution est augmentée d'un *ce* entre les deux termes (*après ce que*, *avant ce que*, *pendant ce que*, etc.), plus rarement antéposé (*ce pendant que*); les formes en *ce* disparaissent, pour la plupart, au xv^e s. (GGHF *id.*, p. 905) — ces formes en *ce* ne font pas partie de nos données, cf. *infra*. La

COMBETTES (2020). « Conjonctions de subordination et locutions conjonctives ».

1: Attestée dès 1219; GGHF *id.* p. 905.

grammaticalisation des locutions conjonctives a principalement lieu en moyen français (GGHF *id.*, p. 906), mais soit toutes ne subissent pas ce processus à la même vitesse, soit ce processus ne les rend pas opaques et est donc réversible; au XVIII^e s., Voltaire peut encore insérer *donc* entre *puis* et *que* (*Puis donc que*; GGHF *id.*, p. 899).²

2: Voltaire (1734), *Lettres philosophiques*, p. 152.

Dans notre corpus, un certain nombre de ces locutions conjonctives sont présentes sous la forme d'un seul mot, et sont donc détectées par Comparalem : *tantque*, *puisque*, *desque*, *ainzque*, etc. — la notion de *mot*, ici, ne renvoie qu'à l'unité tokénisée telle quelle dans le corpus. Ces locutions, qu'elles soient déjà grammaticalisées ou non, peuvent donc être prises en compte dans notre approche; nous parlons de constructions en *base + -que* pour les désigner. Elle fournissent une importante source de données complémentaires, dans la mesure où les conjonctions simples sujettes à un défaut de voyelle finale se limitent à « come » et « que », et où ces deux conjonctions simples sont complexes à traiter par ailleurs — les deux sections suivantes abordent la complexité du cas de « que » et « come ».

Ainsi que présenté en table 10.1, les seules conjonctions simples (non composées) disposant d'une fréquence relativement élevée (ici $F > 200$)³ sont « et », « que », « come », « car » et « mais » et, sur ces cinq conjonctions, seules « que » et « come » connaissent des alternances $\langle e \rangle / \emptyset$. Comme, par ailleurs, l'alternance *come/com* n'est pas toujours considérée comme phonologique et/ou ne correspond pas à un défaut de $\langle e \rangle$ (cf. sec. 10.1.3), les seules conjonctions à connaître un défaut canonique sont « que » et les locutions conjonctives en *base + -que*. Ne pas considérer ces locutions reviendrait donc à ne travailler que sur *que* isolé, ce qui appauvrirait par trop les données.

3: Cette fréquence sert uniquement à exposer la situation générale des conjonctions, et aucun seuil de ce type n'est retenu dans notre traitement des données.

TABLE 10.1 – Lemmes de conjonctions dont la fréquence est supérieure à 200 occ. La correspondance entre lemmes et nombre d'occurrences a été reconstruite après prise en considération des multiples étiquettes de lemmes.

Lemme	Nb. occ.	$\langle e, V \rangle / \emptyset$
<i>et</i>	136 130	Non
<i>que</i> ^a	122 133	Oui
<i>come</i>	13 751	Oui
<i>car</i>	11 163	Non
<i>mais</i>	7646	Non
<i>tantque</i>	1452	Oui
<i>puisque</i>	836	Oui
<i>ainzque</i>	658	Oui
<i>desque</i>	658	Oui
<i>maisque</i>	266	Oui
<i>aincoisque</i>	258	Oui
<i>sique</i>	208	Oui

a. Dont *qui* et occurrences non conjonctives, cf. sec. 10.1.2.

10.1.2 *Que*, une question de nature

Le statut de « que » est complexe : le mot peut être pronom (10.1), conjonction (10.2), adverbe (10.3), et peut parfois assurer plusieurs de ces rôles en même temps (MOIGNET 2002 [1976], p. 272), comme dans les propositions comparatives subordonnées où « que » est à la fois adverbe

MOIGNET (2002 [1976]). *Grammaire de l'ancien français*.

de degré et conjonction de subordination (10.4) — ce qui peut donner lieu à des emplois dédoublés, avec ou sans *ce* interjeté (10.5 - 10.6).

(10.1) *dis blanches mules fist amener marsilies
que li tramist li reis de suatilie*

« Marsile fit amener dix mules blanches
que lui avait envoyées le roi de Süatile »
(*Roland*, v. 89-90, trad. SHORT 1990)

(10.2) *jo vos defend que n i adeist nuls hom*

« je vous défends que personne y touche »
(*Roland*, v. 2438, trad. *id.*)

(10.3) *plus en abat que jo ne vos sai dire*

« [il] en fait tomber plus que je ne sais vous dire »
(*Roland*, v. 2339, trad. *id.*)⁴

4: Cité par MOIGNET 2002 [1976], p. 272.

(10.4) *asez est melz qu il i perdent le chefs*

que nus perduns l onur ne la deintet
« Il vaut bien mieux qu'ils y perdent leur tête
plutôt que nous, nous perdions terres et bien »
(*Roland*, v. 44, trad. *id.*)⁵

5: *Idem.*

(10.5) *Mieuz est il qu'il le giete delez la fontaine si que les bestes del bois le
mangeront, que ce qu'il l'ocie de s'espee.*

« Il vaut mieux qu'il l'abandonne près de la fontaine et que les
bêtes de la forêt le mange, plutôt qu'il ne le tue avec son épée »
(*Le Roman de Tristan en prose*, trad. pers.)⁶

6: *Idem.*

(10.6) *Meuz vaut il que en leauté se teigne sanz Yselt avoir, que qu'il eüst
Yseult en sa saisine a faire tote sa volenté et fust renomez a deleal.*

« Il vaut mieux qu'il demeure loyal sans avoir Yseult, plutôt qu'il
n'ait Yseult en sa possession, qu'elle soit à sa merci et qu'il soit
qualifié de déloyal. »
(*Le Roman de Tristan en prose*, trad. pers.)⁷

7: *Idem.*

L'annotation du corpus ne différencie pas ces emplois : l'étiquette numérique de Dees est unique (deespos = "600"), son explicitation est toujours ambiguë (pos = "PROCON") et l'annotation de Tree Tagger est strictement équivalente (taggerpos = "PROCON"). Cette étiquette PROCON semble correspondre à la fois au statut de pronom relatif (donc conjonctif, nominalisant la proposition qu'il domine) et au statut de conjonction pure ; en (10.2), *que* conjonction pure est annoté pos="PROCON", comme *que* pronominal en (10.1). Il est donc impossible, dans les données du NCA3, d'utiliser l'annotation pour différencier « que » selon qu'il est pronom inanimé, adverbe ou conjonction.

De plus, il n'est pas non plus possible de séparer « que » et « qui », dans la mesure où « qui » s'élide également (MARCELLO-NIZIA 2020d, GGHF, p. 733), où leur annotation se résume dans les deux cas au même attribut PROCON et où les formes élidées des deux termes sont strictement homographes (*qu*, *k*, etc.) et ne sont jamais lemmatisées qui par Tree Tagger. Puisque les formes élidées du pronom relatif animé *qui* sont impossibles à différencier des formes élidées de *que*, pronom relatif inanimé ou conjonction, il est impossible d'interroger les défauts de <e> de *que* seul. Aussi, nous avons décidé d'interroger l'ensemble de ces formes, sous le vocable de « élément *que* », indépendamment de leur statut morphosyntaxique. Par cette expression d'élément « *que* », nous désignons toutes les occurrences de « que » et « qui », élidées ou non,

MARCELLO-NIZIA (2020). « Les relatifs, interrogatifs, exclamatifs : déterminants, pronoms et adverbes ».

pronominales ou purement conjonctives. La méthode qui permet cette interrogation, relativement simple, est détaillée en sec. 10.1.4.

10.1.3 Come, existence puis disparition d'une distribution syntaxique

Le cas du mot « come » est encore plus trouble que celui de l'élément *que*, mais pour des raisons différentes. Il y a, en premier lieu, un problème d'étymologie et d'évolution phonétique, qui semble échapper à la majeure partie de la bibliographie. Pour la plupart des sources, le mot provient du latin QUŌMODŌ par l'intermédiaire de la forme *como* attestée à Pompéi.^{8,9} Si cette forme *como* permet bien de prédire les formes sans <e> (*cum*, *com*, etc.), elle ne peut générer les formes avec <e> (*cume*, *come*, etc.), ce qui conduit É. BOURCIEZ (1955, p. 17) à analyser ce <e> comme une épenthèse dont il ne spécifie pas l'origine. D'autres préfèrent voir l'origine de ce <e> dans une locution latine QUŌMODŌ ET (FEW) ou une locution directement proto-romane *com[o] e[t]* (VISING 1895, p. 122-123; POPE 1966 [1934], p. 119), sans relever la difficulté à faire évoluer ce ET en <e> pour schwa. Dans le REW, la situation est moins tranchée, *com(e)* étant classé sous l'entrée principale QUŌMODŌ tandis que la sous-entrée QUŌMODŌ ET est réservée à des formes comme l'italien *come*.

10.1.3.1 Pertinence d'une distribution syntaxique

Par ailleurs, les dictionnaires et grammaires divergent dans leur traitement de la variation <e>/Ø de *com(e)*, qui est souvent présentée implicitement comme une simple variation graphique, et plus rarement comme une variation d'origine syntaxique. Ainsi, parmi les sources qui ne font pas de distinction entre les deux formes, l'AND classe les deux types sous l'entrée « cum », tandis que c'est l'inverse dans le DMF, où *com* apparaît comme variante graphique sous l'entrée « comme ». La version préliminaire du DEAF dispose également d'une seule entrée (DEAFpré : « come¹ »), qui regroupe les variantes *com* et *come*, tandis que le GD « com » renvoie à l'entrée « come », laquelle enregistre *com* en tant que simple variante. Ces ouvrages traitent donc des deux types, *com* et *come*, comme s'il s'agissait d'un seul et même 'mot' ; c'est aussi l'optique de MÉNARD (1988 [1973], 1997)¹⁰, dont la syntaxe (1988) mentionne presque systématiquement *com(e)*¹¹ ; de même, la grammaire de MOIGNET (2002 [1976]) cite presque systématiquement « *com(e)* » et ne comporte, en plein texte, de mention « com » ou « come » que rarement, et sans commentaire sur le fait que le type monosyllabique ou bisyllabique soit préféré dans tel ou tel usage. De manière analogue, la GGHF traite des formes du type *com* sans distinction avec celles du type *come*.¹² Plusieurs grammaires vont jusqu'à proposer des dérivations diachroniques indifférentes au statut du <e> de *com(e)* : ZINK (2013 [1986], p. 243) donne « *quomo* (*do*) > *como* > *com(e)* »,¹³ BRITAIN (1900, p. 35-36) donne « **quomo* (= *quomodo*) > *comme* »,¹⁴ PIERRET (1994, p. 142) écrit « *quōmōdo* > **quōmōdo* > *comme* » (*id.* p. 355) et NYROP (1899, p. 390) admet une chute de *o* dans « *quomodo* > *comme* », sans autre remarque sur l'origine du <e> (et sans citer la forme *como* des inscriptions pompéiennes).

8: TLL « **quōmōdo* », p. 1287, l. 45-49 : *como*[*do*], *como*, *como*[*d*].

9: Indépendamment de la difficulté à faire évoluer QUŌMODO en *como*, nous acceptons ce dernier comme origine des formes monosyllabiques, comme le fait la majeure partie de la bibliographie.

É. BOURCIEZ (1955). *Précis historique de phonétique française*.

VISING (1895). « *Quomodo* in den romanischen Sprachen ».

10: MÉNARD (1997, p. 267) : « [Les différentes valeurs] étaient impliquées, mélangées, amalgamées en ancien français **dans le même terme *com(me)*** » ; la graisse d'emphase est de notre fait.

11: *Com* sans -e n'est cité que pour un point, pour la subordination inverse (MÉNARD 1988 [1973], p. 334) sans remarque spécifique sur le fait que le mot ne puisse être bisyllabique ici.

12: P.ex. p. 1329, 1373-1375, 1422-1423.

13: Zink traite ici du devenir du /k^w/ initial et non du -e final.

BRITAIN (1900). *Historical Primer of French Phonetics and Inflection*.

14: Tout comme Zink, elle pose cette évolution à propos d'un autre point, ici du développement sans diphtongaison des clitiques.

PIERRET (1994). *Phonétique historique du français, et Notions de phonétique générale*.

NYROP (1899). *Grammaire historique de la langue française*.

Au contraire, le FEW, le TLFi et POPE (1966 [1934], p. 119) notent non seulement une origine distincte des formes avec et sans <e>, mais notent également que leur emploi syntaxique est différencié, les formes bisyllabiques n'étant dans un premier temps utilisées qu'en comparaison elliptique (c'est-à-dire sans que le verbe de la comparante soit exprimé). Cette distinction semble provenir, notamment par l'intermédiaire d'une remarque de SUCHIER (1879, p. 69, rq. 9f) et d'un article de VISING (1895) mentionnés par le TL, de l'édition de la *Chanson de Roland* de GAUTIER (1872). Celui-ci note, dans son glossaire :

CUM. Conjonction. Ce mot, suivant nous, ne vient pas TOUJOURS de *quomodo*, ainsi que l'affirment Littré et Brachet. Il vient tantôt de *quomodo* : *Issi seit CUM vos pleist*¹⁵, et tantôt de *quum* : *CUM je serai à Loün en ma chambre*,¹⁶ etc. En ce dernier cas, l'étymologie *quomodo* nous paraît impossible. = D'où il suit que CUM a deux sens : 1° « Quand, lorsque, dès que », et 2° « De la façon que, ainsi que ». = Il convient d'ajouter que CUM se distingue fort nettement de *cume* dans le texte de la Bodléienne.¹⁷ CUME (qui a seulement le sens de *quomodo*), ne s'emploie qu'avec des substantifs et adjectifs (*CUME fols... Neirs CUME peis*)¹⁸ ; tandis que CUM s'emploie PRESQUE toujours avec un verbe : *Faites la guere CUM vos l'avez enprise*,¹⁹ etc. [...] (GAUTIER 1872, t. II, p. 310; nous reproduisons la typographie mais les notes sont de notre fait.)

Si la citation de Gautier par Suchier peut paraître anecdotique (il ne fait que remarquer que l'un de scribes des sermons rimés qu'il édite respecte le même usage), Vising va plus loin, en reprenant le compte des *cum(e)* de Roland. Il aboutit à la conclusion que la règle relevée par Gautier est juste : sur les 65 occurrences qu'il relève, 37 *cum* introduisent bien une proposition dont le verbe est exprimé, et 21 *cume* introduisent bien une comparaison elliptique. Il n'a donc que 7 contre-exemples, et il propose d'en corriger 3 pour des raisons métriques : en (10.7 - 10.9), il repère que le vers est hypermétrique est que c'est *cum* qui est attendu.

- (10.7) *Cume fist a tei le bastun devant Carle*²⁰
« [le gant me tomberait à terre] comme le baton à toi devant Charles »
(*Roland*, v. 765, trad. SHORT 1990)
- (10.8) *Fiert de l'espier tant cume hanste li duret*²¹
« [il] frappe de l'épieu tant que la hampe reste entière »
(*Roland*, v. 1323, trad. *id.*)
- (10.9) NCA *oliver frere cumment le purrum nus faire*
BFM *Oliver, frere, cum le purrum nus faire?*²²
Dipl. *U luer frere cū, mēt le p^arū nus faire.*
« Olivier, frère, comment pourrons-nous faire ? »
(*Roland* : rolandox (NCA), roland (BFM), v. 1698, trad. *id.*)

SUCHIER (1879). *Reimpredigt*.

VISING (1895). « Quomodo in den romani-schen Sprachen ».

GAUTIER (1872). *La Chanson de Roland*.

15: V. 606 : « Qu'il en soit comme il vous plaît! », trad. SHORT (1990).

16: V. 2910 : « Quand je serai à Laon, dans mon domaine. », trad. *id.*

17: Bibliothèque qui consigne le manuscrit de référence de *Roland*, dit « manuscrit d'Oxford », qui est également repris par rolandox (NCA : SEGRE 1971) et roland (BFM : MOIGNET 1972a).

18: V. 1474 : *Issi est neirs cume peiz ki est demise*, « la poix fondue n'est pas plus noire que lui », trad. *id.*

19: V. 210 : « Faites donc la guerre comme vous l'avez entreprise », trad. *id.*

20: SHORT (2007) corrige : *cum*.

21: *Id.*

22: *Idem* GAUTIER (1872) et SHORT (1990)



FIGURE 10.1 – *Roland*, ms. d'Oxford, fol. 31r, l. 15 (v. 1698). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

Si nous nous accordons avec la plausibilité de la correction du premier vers cité (10.7), nous notons que, dans le second vers (10.8), la correction n'est nécessaire que si l'on considère que le 'h aspiré' interdit systématiquement l'élision, or nous avons vu que ce n'était pas systématiquement le cas dans *Roland* (cf. 8.2.2.1 et 6.2.3). Enfin, dans le troisième exemple (10.9), ce n'est pas *cume* que porte le manuscrit, mais une abréviation en partie surscrite (fig. 10.1), que SEGRE (1971, NCA) rend par *comment*, ce qui rend effectivement le vers hypermétrique. Ici, la plupart des éditeurs semblent préférer ignorer le trait vertical abrégatif et les signes surscrits, éditant *cum* non par correction graphique mais en déclarant caduque l'abréviation - celle-ci, en effet, rend le vers hypermétrique et a visiblement été ajoutée après la copie du vers.

Vising note également que deux contre-exemples où *cume* introduit une proposition dont le verbe est exprimé se trouvent devant initiale vocalique, sous-entendant que cela autorise la correction de *cume* en *cum* (10.10-10.11). Sur ce point, s'il est vrai que la métrique ne s'oppose pas à cette correction, il faut objecter qu'elle n'indique absolument pas la nécessité de l'élision non plus et que le scribe aurait tout aussi bien pu choisir la forme graphique correspondant à l'introduction d'une comparaison non elliptique, c'est-à-dire *cum*.

(10.10) *e il li ad cum chevaler mustree*
« et il la lui fait voir en vrai chevalier »
(*Roland*, v. 1369, trad. SHORT 1990)

(10.11) *e il les pluret cum chevaler gentill*
« et il les pleure en noble chevalier »
(*Roland*, v. 1854, trad. SHORT 1990)

En comparant le relevé d'occurrences de Vising avec ce que nous trouvions dans le NCA, nous avons remarqué qu'il manquait 5 occurrences dans le relevé de Vising (qui, fort heureusement, ne sont pas des contre-exemples), et qu'il s'était apparemment trompé sur le classement d'une occurrence.²³ Dans l'ensemble du texte, nous avons trouvé 46 graphies *cum* dont 4 introduisent une comparaison elliptique; ce sont les deux contre-exemples notés tels quels par Vising (10.10 - 10.11) et les deux cas devant initiale vocalique où il suggère que la métrique autorise la correction de *cum* en *cume* (10.12 - 10.13).

(10.12) *karles se dort cum hume traveillét*²⁴
« Charles s'endort comme un homme accablé »
(*Roland*, v. 2525, trad. SHORT 1990)

(10.13) *vait le ferir cum hume vertudable*
« [il] va le frapper en homme très vigoureux »
(*Roland*, v. 3426, trad. SHORT 1990)

Si ces quatre contre-exemples semblent marginaux vis-à-vis de l'ensemble des 46 occurrences *cum* du texte, nous notons néanmoins que *cum* ne se trouve directement devant un nom que dans 8 occurrences; ces quatre contre-exemples représentent donc la moitié des occurrences *cum* + Nom. Quant aux graphies *cume*, nous en avons trouvé 24 dans *roLandoX*, dont trois introduisent une proposition dont le verbe est élidé. Vising note également trois contre-exemples; les deux premiers sont corrects (10.10-10.11 *supra*), mais le troisième n'est pas un contre-exemple — Vising fait erreur sur le classement du v. 2414 comme mentionné *supra* (note 23). Il lui

23: Vising classe le v. 2414 comme *cume* en comparaison elliptique, alors que les éditions et le ms. donnent *cum hom est irét* (« comme un homme désespéré », trad. SHORT 1990). L'exemple ne contredit donc en rien son analyse, puisque *cum* y introduit une proposition non elliptique.

24: En de rares occasions, comme pour les participes passés passifs de *Roland*, le NCA conserve des diacritiques éditoriaux indiquant que le mot est oxyton.

manque donc un contre-exemple, qui se trouve au v. 3698 (10.14). Dans ce vers, le premier hémistiche du vers est hypométrique indépendamment de *cume*; SHORT insère un *si* initial pour former la locution *si cume* et rétablir l'isosyllabisme. Absent du relevé de Vising, ce vers constitue bien un contre-exemple, *cume* introduisant une proposition dont le verbe est exprimé (*est*), et la forme *cume* étant sans ambiguïté sur le ms. (fig. 10.2).

- (10.14) NCA *cume il est en sun paleis halcur*²⁵
 SHORT *Si cume il est en son paleis halçur*
 Dipl. *C ume il .é. en sun paleis halcur.*
 « Sitôt qu'il est dans son palais principal »
 (*Roland* : ro landox (NCA), ro land (BFM), SHORT 1990 v. 1698,
 trad. SHORT 1990)

25: *Idem* GAUTIER (1872) et MOIGNET (1972a, BFM); Gautier édite *haltur* au lieu de *harçur*.

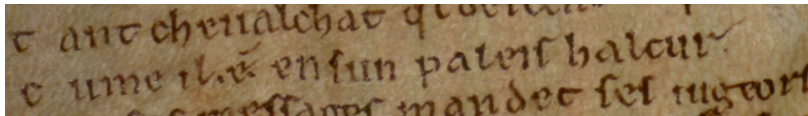


FIGURE 10.2 – *Roland*, ms. d'Oxford, fol. 66v, l. 26 (v. 3698). Source : Bodleian Library MS. Digby 23, Part 2, <https://digital.bodleian.ox.ac.uk/inquire/p/79097275-ef1d-4107-85d3-e8402120f365>.

Si l'on s'abstient de toute correction métrique, les contre-exemples représentent 10% des occurrences, aussi la distribution complémentaire des formes n'est pas aussi claire que ce qu'avance Vising. En appliquant ses corrections métriques, les contre-exemples ne représentent plus que 3% des occurrences. Surtout, c'est le contexte *cum* + Nom qui semble être sujet à variation, comme si celui-ci ne permettait pas à *cum* de percevoir si le nom était sujet d'un verbe exprimé ou non. Hormis peut-être devant nom, *Roland* témoigne donc d'une distribution presque complémentaire, réservant majoritairement *cume* à l'introduction d'une proposition sans verbe exprimé.

10.1.3.2 Affaïssement de la distribution syntaxique

Nous avons voulu vérifier la diachronie et la diatopie avancées par Vising, qui relève que la distinction *cum* ~ *cume* se maintient à l'Est et au Nord du domaine, alors qu'elle s'affaiblit en anglo-normand plus rapidement. Notre but est notamment d'identifier, ne serait-ce qu'approximativement, le moment où cette distinction entre *cum* et *cume* cesse d'être valide. Le corpus n'étant pas annoté en syntaxe, nous nous sommes concentré sur les attributs de partie du discours suivant *cum* et *cume* dans *Roland*. Sur la base de ces attributs, nous ne pouvons interroger les contextes attribués à *cume* (comparaison elliptique), les valeurs de pos du mot suivant étant également possibles dans des propositions non elliptiques, mais nous pouvons mesurer les taux de formes *cume* dans deux contextes qui ne peuvent être elliptiques et qui sont donc *a priori* ceux de *cum* : devant pronom personnel sujet et devant verbe conjugué. Ces contextes ont été vérifiés manuellement pour confirmer qu'il s'agit bien de propositions dont le verbe est exprimé. Le nombre de configurations *cum*(e) + VER (988 occ.) est bien moins élevé que celui de configurations *cum*(e) + PROPERSONS (3756 occ.), mais, dans les deux cas, l'opposition semble commencer à se neutraliser dès après *Roland* : si certains textes maintiennent une distinction plus ou moins systématique, d'autres ont moins de 50% de

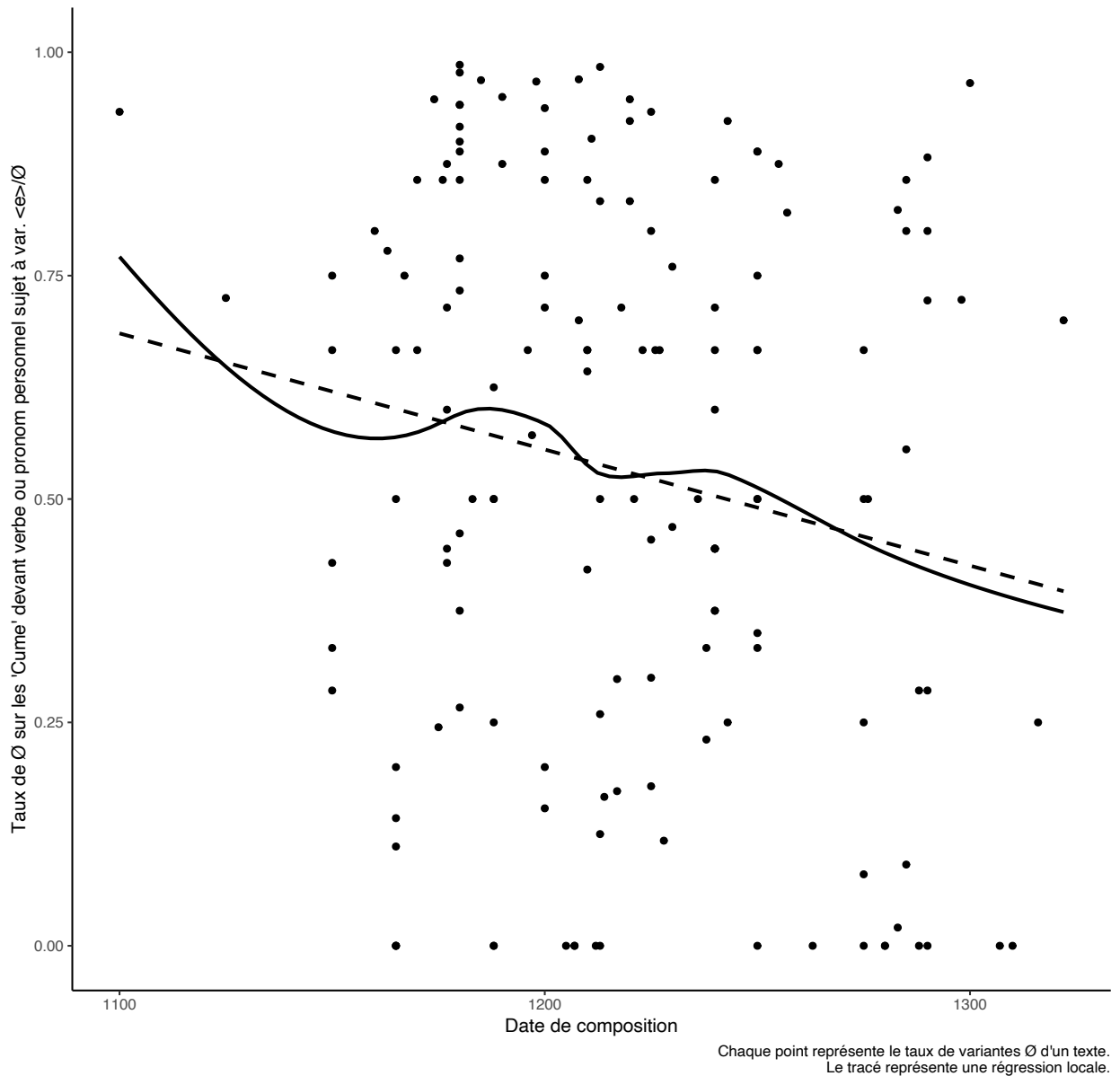


FIGURE 10.3 – Distribution diachronique des taux de *come* bisyllabiques devant pronom personnel sujet et verbe conjugué.

cum devant proposition dont le verbe est exprimé, et ce dès le milieu du XII^e s., tel que représenté en fig. 10.3. On voit donc, sur ce graphique, se dessiner la neutralisation progressive de l'opposition *cum* ~ *cume*, opposition que l'on peut estimer caduque dès la seconde moitié du XII^e s.

Certes, le <e> des formes bisyllabiques du type *cume*, *come* n'est pas étymologique, et toutes les données de cette thèse, jusqu'à présent, respectent ce critère de manière à exclure d'éventuelles épenthèses ponctuelles. Mais, ici, une fois la neutralisation avérée, il ne fait guère de doute qu'il n'y a plus qu'un seul mot, sujet à une variation <e>/Ø qui ne tient pas à des épenthèses ponctuelles — dès *Roland*, *cume* est attesté. Dans ce qui suit, nous intégrons donc ce mot dans l'analyse, de manière

à compenser la faiblesse des données, qui se résumeraient, sans cela, aux conjonctions en *base* + *-que* et aux *que* isolés. Comme nous traitons séparément les données des trois types de conjonctions envisagées, cela n'aboutit pas à une fusion des données et le lecteur qui voudrait refuser à *come* sa présence dans nos données pourrait aisément l'isoler.

10.1.4 Traitement des données

Comparalem a détecté 42 lemmes de conjonctions dotés d'une potentielle variation <e>/Ø; ces lemmes correspondent à des locutions conjonctives dont le dernier terme est *-que* (et qui sont écrites en un seul mot dans le corpus) et à la conjonction « come ». Pour les conjonctions en *-que*, l'étiquette de lemme choisie par Tree Tagger correspond presque toujours à la base sur laquelle s'ajoute « que » et n'est donc pas appropriée — les conjonctions lemmatisées ainsi sont toutes du type *ainsique*, celles lemmatisées combien du type *combienque*, tandis que les occurrences du type *deceque* sont lemmatisées de et celles du type *decique* deci. Ces étiquettes sont surnuméraires dans de nombreux cas, avec par exemple *dementiers* pour le type *dementiersque* et *dementre* pour le type *dementreque*, alors qu'il s'agit d'une simple variation de la même base.²⁶ Sur ces 42 étiquettes de lemmes, nous en avons éliminées 3, qui regroupent sous la même étiquette des lemmes différents (p.ex. types *atantque* et *afinque* regroupés sous *a*, ou *mais* et *maisque* regroupés sous *mais*). Une fois l'opération de relemmatisation effectuée (cf. sec. 10.1.4.2 *infra*), les 39 étiquettes de Tree Tagger sont réduites à 31 étiquettes, auxquelles s'ajoute l'élément *que*; nous travaillons donc ici sur 32 lemmes. Enfin, nous avons ponctuellement écarté 17 occurrences : 9 comportent un <a> final à la place d'un <e> final (8 occ. *coma* dans *poit* et 1 occ. *quoma* dans *alia*), et 8 correspondent à autre lemme que celui qui leur est attribué alors que les autres attributions du lemme sont correctes (p.ex. 1 occ. *de* pour lemma = "de" alors que les autres occurrences lemma = "de" & pos = "CON" correspondent au type *deceque*). Enfin, nous avons contrôlé dans TXM, par des requêtes d'index sur les lemmes, que ceux-ci ne correspondaient qu'aux formes désirées.

Étant donnés les éléments développés *supra* concernant l'élément *que* et le mot *com(e)*, nous avons adopté un processus de collecte des formes qui est mixte : les formes sont collectées normalement pour les 30 lemmes en *base* + *-que*, tandis qu'elles sont collectées par des requêtes spécifiques pour l'élément *que* et le mot *com(e)*. Nous détaillons ces requêtes avant de lister les opérations de relemmatisation effectuées.

10.1.4.1 Requêtes pour l'élément *que* et pour *com(e)*

Les requêtes servant à capturer les conjonctions suivent donc le format habituel de nos requêtes CQP, telles que présentées en sec. 3.1.1. Pour capturer l'ensemble des formes de l'élément *que*, ces requêtes n'étaient pas appropriées, étant donné que nous cherchons ici à opposer les formes en <e> et <i> aux formes sans voyelle finale. Par ailleurs, Tree Tagger propose 4 lemmes différents pour l'élément *que*, à savoir *que2_que3_que* (F = 52 298), qui (F = 39 293), *que1_que2_que3_que4_que2* (F = 5954) et *que1_que2_que3_que4* (F = 599),²⁷ et dont seul qui dispose d'une

26: LAND (« dementers ») et le DMF (« dementiers ») enregistrent les deux formes comme une simple variation graphique; GD dispose de deux entrées mais sans nuance sémantique patente. Le FEW dispose de deux sous-entrées (sous « dum interim »), et indique que la variation est au moins en partie dialectale: « Unklar its das verhältnis zwischen [type demientre] und [type dementiers]; diese letztern formen sind im wesentlichen auf die nördlicheu idiom beschränkt » (*La relation entre le type demientre et le type dementiers n'est pas claire; les formes du second type se trouvent essentiellement limitées aux dialectes septentrionaux; trad. pers.*).

poit : ANON. (1217*). *Chronique traduite de l'Historia Karoli Magni et Rotholandi*. Dees : Vendée, Deux-Sèvres; compo. saint., ca 1217; ms. saint., ca 1238; MS1 (AURACHER 1877); DEAF : Turpin¹A.

alia : ANON. (1185*). *Roman d'Alexandre*. Ms. A; Dees : Vendée, Deux-Sèvres; compo. pic. et occ. orient., ca 1185; ms. poit., ca 1225; MS (ANON. 1937); DEAF : AlexArsL.

27: Ces fréquences sont celles de l'ensemble du corpus; les données sur lesquelles nous travaillons sont moins amples du fait des restrictions contextuelles de Comparalem et de l'exclusion des textes trop tardifs ou non datés.

attribution qui nous a semblé cohérente (formes *qui*, jamais élidées). Nous avons donc établi l'ensemble des formes que pouvait prendre l'élément « que », telles que représentées en table marginale 10.2, et les avons collectées avec les requêtes en (10.15) pour les formes à finale vocalique et (10.16) pour les formes en défaut vocalique ; la spécification du contexte n'est pas indiquée ici (cf. sec. 3.1.1). Nous avons contrôlé que l'ensemble des formes répondant à ces requêtes ne correspond qu'aux quatre lemmes mentionnés, c'est-à-dire qu'aucun autre lemme ne dispose d'occurrences homographes de l'une de ces formes qui apparaissent annotées pos = "PROCON".

(10.15) [word = "(q|k)u?(e|i|y)" & lemma = "qu(e|i).*" & pos = "PROCON.*"]

(10.16) [word = "(qu|k|q)" & lemma = "qu(e|i).*" & pos = "PROCON.*"]

De même, pour *com(e)*, Tree Tagger propose trois lemmes différents, *come1* (F = 9383), *come1_come* (F = 3433) et *come* (F = 921), dont la distribution ne correspond pas aux alternances <e>/∅ et ne semble pas non plus correspondre à une position syntaxique spécifique. Comme pour l'élément *que*, nous avons rédigé des requêtes capturant l'ensemble des formes du mot *com(e)*, qui sont données en (10.17) et (10.18). L'ensemble des formes répondant à ces requêtes ne correspond qu'aux trois lemmes mentionnés, et est indiqué en table marginale 10.3. Lors de la saisie des contextes subséquents dans les requêtes, nous avons inclu les différentes formes du pronom sujet *il* (d'ordinaire exclu puisque commençant par un <i>), dans la mesure où ces requêtes servent notamment à documenter la configuration *come* + PROPersCS.

(10.17) [pos = "CON.*" & word =
"(c|k|qu?)(o|u)(m|n)(m|n)?e"]

(10.18) [pos = "CON.*" & word = "(c|k|qu?)(o|u)(m|n)(m|n)?"]

10.1.4.2 Relemmatisation

Comme exposé dans la section précédente, les lemmes de Tree Tagger pour les conjonctions correspondent, pour une large majorité des locutions conjonctives, au lemme de la base, c'est-à-dire le plus souvent à l'adverbe auquel est adjoind *-que*. À cela s'ajoute le fait que la conjonction « come » dispose de trois étiquettes (*come*, *come1* et *come1_come*), et que l'élément *que* correspond aux quatre étiquettes mentionnées en section précédente. Aussi, nous avons procédé à une relemmatisation générale des données, selon les correspondances affichées en (10.19). Dans le cas des correspondances bijectives (un lemme de Tree Tagger = 1 étiquette re lemma), l'opération de relemmatisation ne change rien à l'analyse, mais évite de travailler sur des étiquettes qui ne correspondent pas aux formes qu'elles recouvrent et simplifie donc la gestion des données. Au demeurant, dans la mesure où tous les lemmes sauf « come » et « que » sont traités ensemble dans ce qui suit, il importe peu de savoir s'il faudrait, par exemple, regrouper *einsique* et *ainzque*.

TABLE 10.2 – Formes et fréquence des formes de l'élément *que* dans le NCA.

forme	F
<i>que</i>	52 298
<i>qui</i>	33 335
<i>qu</i>	20 724
<i>ke</i>	5954
<i>ki</i>	5769
<i>k</i>	2443
<i>qi</i>	816
<i>q</i>	599
<i>qi</i>	138
<i>ky</i>	20
<i>kui</i>	14
<i>quy</i>	13
<i>qy</i>	4

TABLE 10.3 – Formes et fréquence des formes de « come » dans le NCA.

forme	F
<i>com</i>	4485
<i>comme</i>	2634
<i>cum</i>	2261
<i>con</i>	1696
<i>come</i>	941
<i>conme</i>	815
<i>cume</i>	768
<i>cumme</i>	104
<i>conm</i>	15
<i>comm</i>	11
<i>cun</i>	5
<i>cunme</i>	1
<i>kom</i>	1

(10.19) Tree Tagger	Nouveau lemme
ainçois	→ ainçoisque_relemma
adonc	→ adontque_relemma
ainsi, en	→ einsique_relemma
ainz	→ ainzque_relemma
après	→ apresque_relemma
avant	→ avantque_relemma
cant2	→ quantque_relemma
combien	→ combienque_relemma
come, come1	→ come_relemma
coment	→ comentque_relemma
de	→ deceque_relemma
deci	→ decique_relemma
dementiers, dementre	→ dementiersque_relemma
des	→ desque_relemma
devant	→ devantque_relemma
endementiers, endementre	→ endementiersque_relemma
entrementres	→ entrementresque_relemma
entrues	→ entresque_relemma
fors	→ forsque_relemma
freci	→ frecique_relemma
lors	→ lorsque_relemma
luès	→ luesque_relemma
maintenant	→ maintenantque_relemma
ou	→ ouque_relemma
por	→ porque_relemma
puis	→ puisque_relemma
que, que1, que2, que3, que4, qui	→ que_relemma
si	→ sique_relemma
tandis	→ tandisque_relemma
tant	→ tantque_relemma
tantost	→ tantostque_relemma
tresque1, tresque1-trosque	→ tresque_relemma

10.2 Présentation des résultats

Étant donnée l'ampleur des opérations de relemmatisation et le fait que nous avons procédé à la fusion de trois jeux de données, il n'est pas approprié de présenter les résultats comme pour les premières catégories du discours, avec une table présentant le nombre de lemmes et de formes, et le nombre de lemmes et de formes sujets à variation. En table 10.4, nous présentons donc les résultats d'une manière analogue à ce que nous avons proposé pour les déterminants et pour les pronoms, cette fois-ci selon que les formes correspondent à une locution en *base + -que*, à la conjonction « come » ou à l'élément *que*. La seconde colonne représente le nombre d'occurrences totales passées dans Comparalem, et les colonnes suivantes représentent le nombres d'occurrences (avec et sans <e,V> finale) ainsi que le taux de formes défectueuses de ces occurrences en fonction de l'initiale du mot suivant. La dernière ligne indique la situation pour l'ensemble des trois sous-catégories détaillées.

Les données présentées en table 10.4, même prises dans leur ensemble (ligne Σ), indiquent déjà un fort effet du contexte sur les taux de défaut.

TABLE 10.4 – Quantité de formes et taux de formes en défaut de <e> pour les conjonctions.

	Nb.	__#C		__#V	
		Nb.	Tx_Ø	Nb.	Tx_Ø
<i>base + -que</i>	2937	1985	0,45%	952	61%
<i>Come</i>	10 039	8252	61%	1787	62%
<i>Que</i> ^a	90 250	68 188	0,11%	22 062	39%
Σ	103 226	74425	7%	24 801	42%

a. Élément *que* : cf. 10.1.2; formes *qui* incluses.

Ainsi que nous allons le détailler dans les sections suivantes, cet effet tient en réalité à une prohibition des défauts devant initiale consonantique pour deux catégories (*base + « que »* et élément *que*), mitigé par une absence d'effet du contexte pour la conjonction « *come* ».

10.2.1 Typologie

Comme suggéré en table 10.4 *supra*, nos trois groupes de conjonctions se comportent différemment. Le critère de la pertinence du contexte subséquent (10.20 a) oppose les conjonctions en *base + -que* et l'élément *que* seul à la conjonction « *come* » : alors que les défauts de <e> de « *come* » ne semblent pas sensibles au contexte (61% __#C, 62% __#V, cf. sec. 10.2.4), les conjonctions en *base + -que* et l'élément *que* seul connaissent beaucoup plus de défauts devant initiale vocalique (*base + -que* : 61%; élément *que* : 39%) que devant initiale consonantique (*base + -que* : 0,45%; élément *que* : 0,11%; cf. sec. 10.2.2 et 10.2.3). Un classement morphologique maintenant le lien entre le suffixe *-que* des locutions conjonctives et l'élément *que* autonome viendrait corroborer ce rapprochement. En revanche, un critère qui serait fondé uniquement sur la valeur des taux de défaut (lorsqu'il y a défaut; 10.20 b) pourrait regrouper les conjonctions en *base + -que* et « *come* », les opposant à l'élément *que* seul : le taux de défaut des conjonctions en *base + -que* devant initiale consonantique (61%) est équivalent au taux de défaut de « *come* » (61% __#C, 62% __#V), tandis que le taux de défaut de l'élément *que* seul est inférieur (39% __#V). Ces entrées typologiques sont reprises en (10.20).

(10.20) Typologie selon les taux de défaut

(a) Selon l'effet du contexte

- .i. Effet fort du contexte : *base + -que*, élément *que* seul
- .ii. Effet nul du contexte : « *come* »

(b) Selon la valeur des taux de défaut²⁸

- .i. 61-61% : *base + -que* (__#V), « *come* » (__#{C, V})
- .ii. 39% : élément *que* seul

28: Uniquement lorsqu'il y a défaut; cette entrée typologique ne considère pas les taux inférieurs à 1%.

Phonologiquement, deux critères peuvent classer ces conjonctions : d'une part, la voyelle susceptible d'être défectueuse est toujours un schwa dans les conjonctions en *base + -que* et « *come* », tandis qu'elle peut être un schwa ou un /i/ pour l'élément *que*, qui regroupe aussi des formes *qui*

(10.21 a). D'autre part, si l'élément *que* est monosyllabique, « come » est bisyllabique ; sur ce critère, le maintien d'un lien morphologique entre le suffixe *que* des conjonctions en *base + -que* et l'élément *que* seul permettrait de les regrouper, en opposition à « come » (10.21 b). Ce regroupement en opposition à « come » pourrait expliquer la sensibilité au contexte : l'élément *que*, seul ou intégré à la fin d'une locution conjonctive, est traité comme monosyllabique et, comme pour les pronoms et articles monosyllabiques, il est sujet à un effet très fort du contexte, tandis que « come », comme le pronom personnel *ele* et l'article *une*, est bisyllabique et n'est pas sujet à un tel effet du contexte.

(10.21) Typologie phonologique

- (a) Selon la voyelle déficiente
 - .i. Uniquement /ə/ : *base + -que*, « come »
 - .ii. /ə, i/ : élément *que* seul
- (b) Selon le nombre de syllabes
 - .i. Monosyllabique : *base + -que*, élément *que* seul
 - .ii. Polysyllabique : « come »

D'emblée, le croisement entre ces quatre typologies permet d'émettre une hypothèse sur l'absence d'homogénéité de ces résultats : *base + -que* se comporte en partie comme l'élément « que » seul (10.20 a et 10.21 b) et en partie comme la conjonction « come » (10.20 b et 10.21 a). Si l'on admet un tel croisement de catégories, on peut admettre plus aisément la mixité des résultats : d'une part l'alternance <e>/Ø des locutions *base + -que* sont contextuelles et leur « -que » est traité comme l'élément *que* pour la sensibilité au contexte — donc comme un monosyllabe ; d'autre part, et contrairement à l'élément *que* seul, les locutions *base + -que* ne connaissent qu'une seule voyelle déficiente, <e> pour /ə/, ce qui est aussi le cas de « come ». Cette seconde différence, *in fine*, peut être ramenée au problème de nature de l'élément *que* : ce n'est peut-être pas la différence <i>/<e> (*qui, que* → *qu*) qui est ici pertinente, puisque *qui* s'élide également, mais bien la structure syntaxique dans laquelle les formes non conjonctives de l'élément *que* s'inscrivent qui fait baisser leur taux de défaut, vis-à-vis des locutions, où *que* est toujours conjonctif.

Dans ce qui suit, nous présentons plus en détail les résultats pour les conjonctions en *base + -que* (sec. 10.2.2), pour l'élément *que* (sec. 10.2.3) puis pour la conjonction « come » (sec. 10.2.4). Nous détaillons l'effet des variables diachroniques et diatopiques pour chacune de ces catégories.

10.2.2 Locutions en *base + -que*

Cette section s'attache à décrire l'alternance <e>/Ø des locutions conjonctives en *base + -que* tokénisées en un seul mot dans le corpus. Elle aborde la distribution contextuelle et diachronique de ces alternances (sec. 10.2.2.1), avant d'aborder la diatopie (sec. 10.2.2.1) et l'intersection entre diatopie et diachronie (10.2.2.3) de cette distribution.

10.2.2.1 Distribution diachronique et contextuelle

Les graphiques de séries temporelles en fig. 10.4.a-d, page suivante, représentent le taux de défaut de <e> des conjonctions en *base* + *-que*, pour chaque texte, avec (fig. 10.4.c,d) et sans (fig. 10.4.a,b) pondération des taux par le nombre d'occurrences de chaque texte. La différence entre les contextes préconsonantiques et prévoicaliques, déjà notée en table 10.4, est confirmée sur la majeure partie de la période : les occurrences défectueuses sont marginales devant initiale consonantique (moy. < 0,01%), alors qu'elles sont largement répandues devant initiale vocalique (moy. = 69%). Les défauts ne sont néanmoins pas systématiques devant initiale vocalique, les taux individuels couvrant l'ensemble de l'échelle des ordonnées, de 0 à 100% ($\sigma = 0,36$; σ pond. = 0,37).

Si les graphiques en fig. 10.4.b,d semblent témoigner d'une tendance à l'augmentation des taux de défauts, la corrélation linéaire est toujours faible ($\rho = 0,12$; ρ pond. = 0,24), et non significative dans la version non pondérée des données ($p = 0,08$). Nous montrerons en sec. 10.2.2.3 que cette tendance faible à la hausse est due à une interférence du sous-corpus anglo-normand, dans la même configuration que pour les adverbes (cf. sec. 7.3.3). Ainsi que représenté sur les boîtes à moustache en fig. 10.4.e, devant initiale vocalique, seul un quart des textes disposent de taux de défauts inférieurs à 50%; aussi, si l'élision n'est pas systématique, elle demeure nettement favorisée dans les trois quarts des textes. Par conséquent, l'effet du contexte est significatif ($p < 0,005$), mais bien moins marqué ($t = 26,48$; t dép. = 25,43) que pour les articles définis *la* ($t = 532$), *le* ($t = 350$) ou la préposition *de* ($t = 86$).

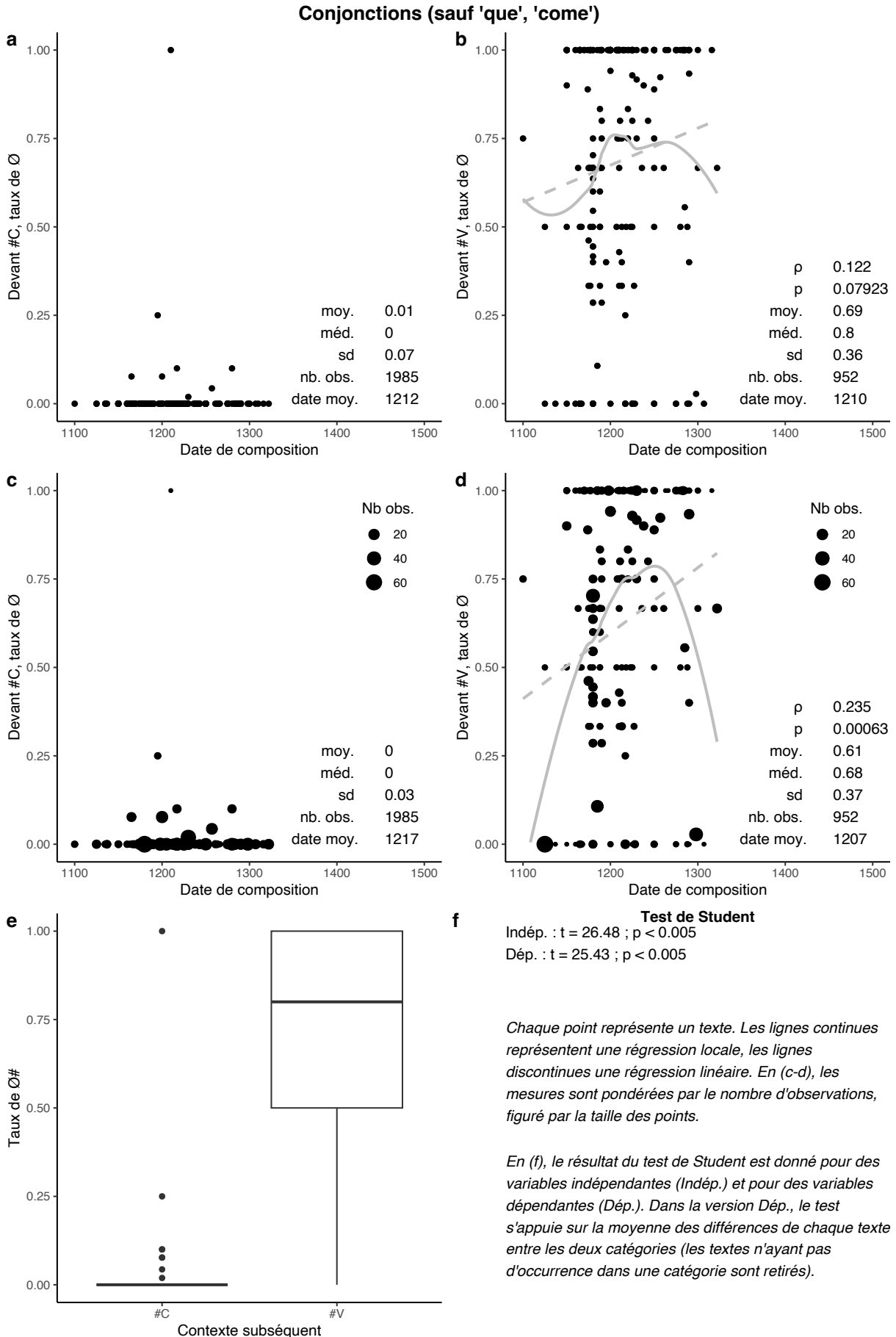


FIGURE 10.4 – Distribution contextuelle des variantes de conjonctions *base* + *-que* en défaut de <e>.

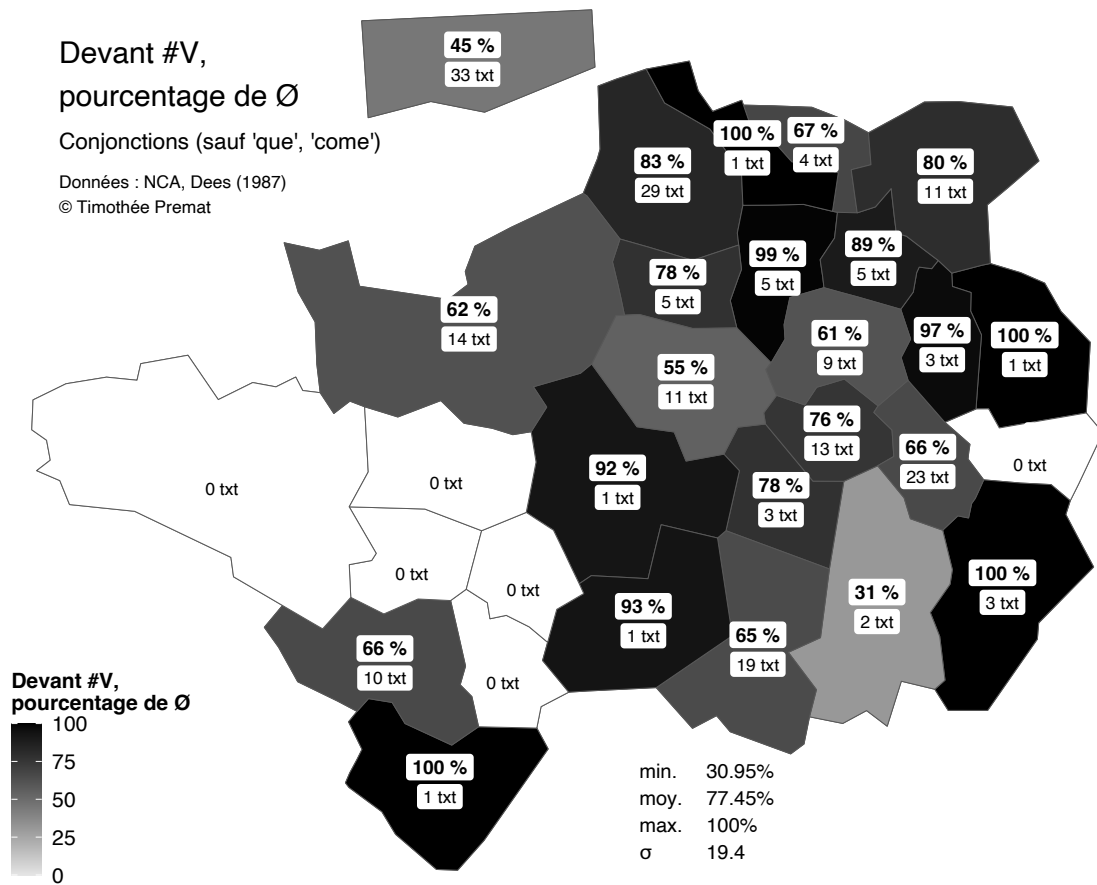


FIGURE 10.5 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, conjonctions en *base* + *-que* suivies d'une initiale vocalique.

10.2.2.2 Distribution diatopique

La distribution diatopique des formes élidées, telle que représentée en fig. 10.5, est peu homogène ($\sigma = 19,4$), avec un minimum régional moyen de 31% (Bourgogne, 2 textes) et un maximum régional moyen de 100% (Franche-Comté, 3 textes; Charente-Maritime, Nord et région Moselle, Meurthe-et-Moselle, 1 texte par région). La moyenne des taux régionaux d'élision s'établit à 77%, soit 8% plus haut que la moyenne des textes en fig. 10.4.b; cela s'explique principalement par le fait que les régions qui ont le plus de textes disposent de taux inférieurs à la moyenne des taux régionaux. À l'exception du Hainaut et de la région Vendée, Deux-Sèvres, ces régions aux taux inférieurs à la moyenne régionale se répartissent sur une diagonale Nord-Ouest - Sud-Est, de l'Angleterre (45%, 33 textes) et la Normandie (62%, 14 textes) à la Bourgogne (31%, 2 textes) et à la région Nièvre, Allier (65%, 19 textes). La diatopie des cas d'élision semble donc suivre un dessin original, l'espace d'oïl étant coupé en deux par une 'vallée' dont les taux sont sensiblement inférieurs à la moyenne non seulement régionale mais également générale (telle qu'en fig. 10.4.b).

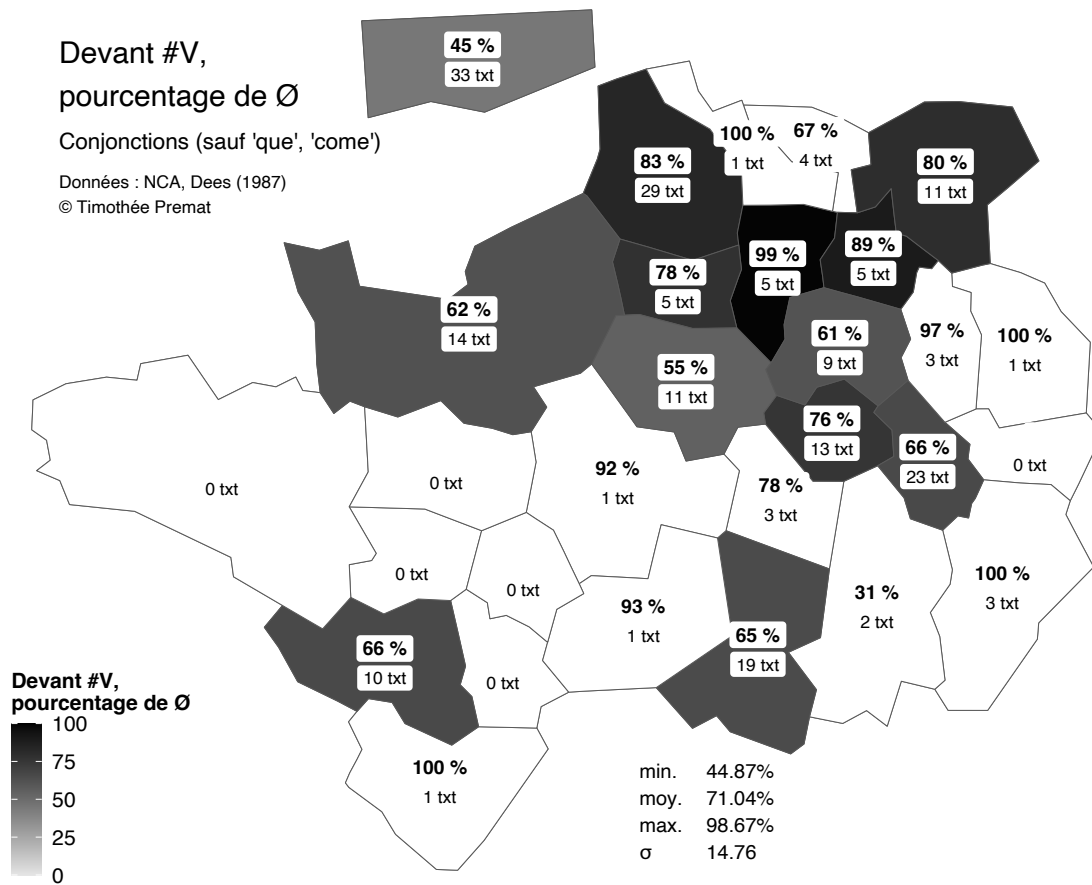


FIGURE 10.6 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, conjonctions en *base* + *-que* suivies d'une initiale vocalique, et à l'omission des régions disposant de moins de cinq textes.

L'amplitude de la variation visible sur les graphiques de séries temporelles est donc en bonne partie une variation diatopique. Il convient néanmoins de noter que la délimitation Sud-Ouest de la 'vallée' est peu fiable, dans la mesure où seule la région Vendée, Deux-Sèvres dispose, au Sud de la diagonale identifiée, de plus d'un texte, et que son taux régional de défauts est faible (66%, 10 textes). Il convient donc d'envisager que la 'vallée' qui se dessine sur nos cartes peut correspondre également à une partition Sud-Ouest *vs* Nord-Est. C'est à peu près l'effet obtenu en fig. 10.6, où toutes les régions disposant de moins de cinq textes sont omises.

10.2.2.3 Intersection de la diatopie et de la diachronie

Cette distribution diatopique ne semble pas dépendre trop fortement de la diachronie. D'une part, la tendance diachronique est faible ($p = 0,12$ en fig. 10.4.b), d'autre part, les cartes divisées par demi-siècle, en fig. 10.7 et 10.8, pages suivantes, témoignent toujours de la présence de cette 'vallée' de taux plus faibles en diagonale du territoire. En effet, si

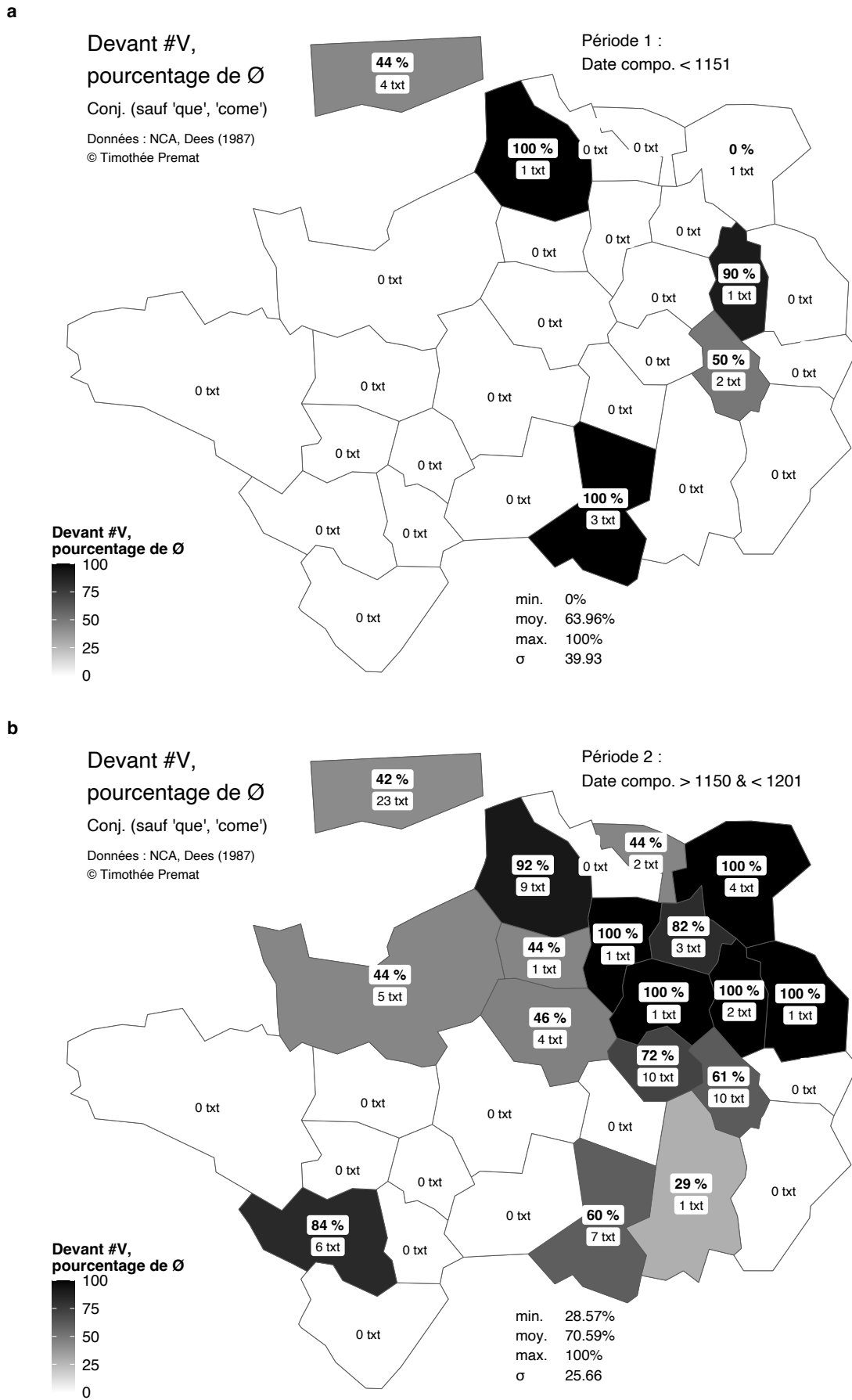
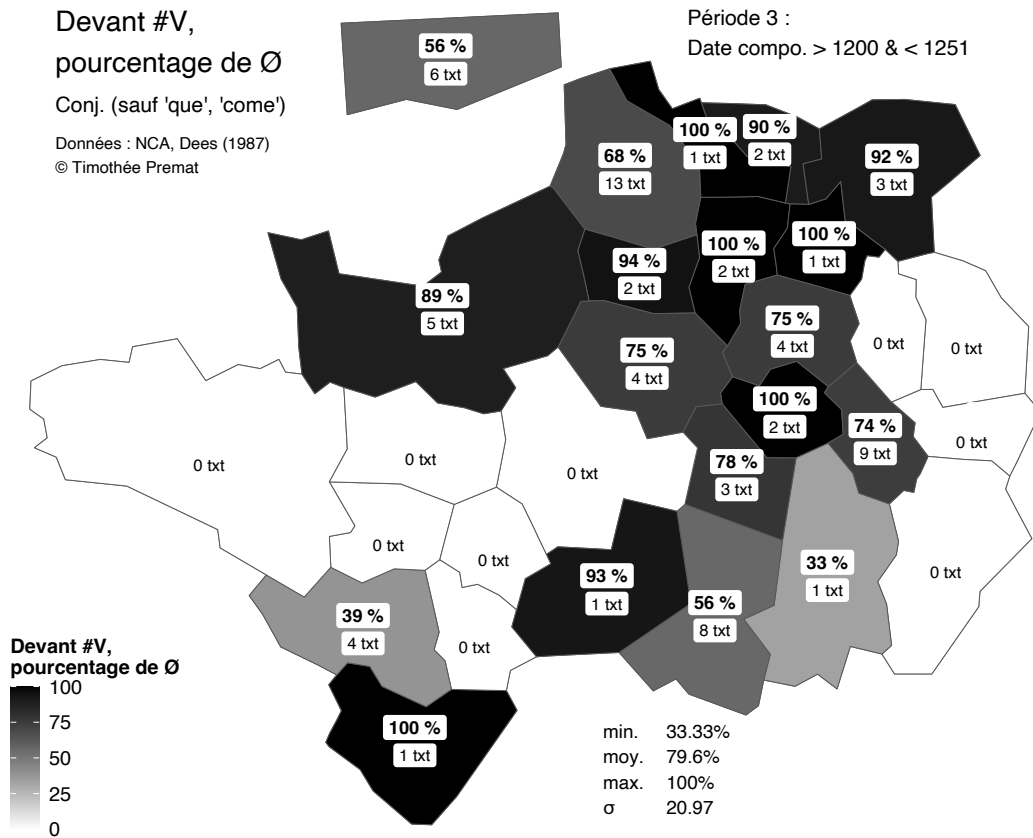


FIGURE 10.7 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1150) et 2 (b : 1151-1200), conjonctions en base + -que.

a



b

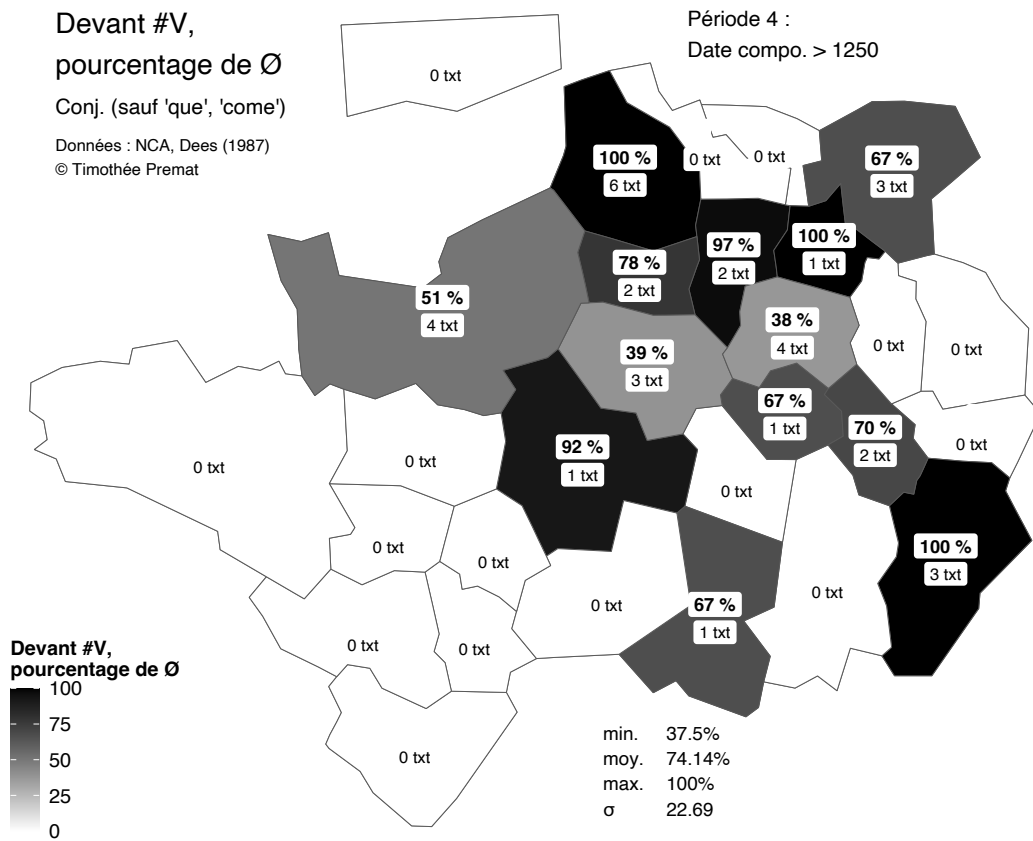


FIGURE 10.8 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), conjonctions en base + -que.

ces cartes montrent une forte fluctuation des taux moyens régionaux, les régions appartenant à cet axe se situent à toutes les époques en-deçà de la moyenne des taux régionaux, à l'exception de quelques cas. Ainsi, dans la seconde moitié du XII^e s. (fig. 10.7.b), l'Aube dépasse légèrement la moyenne des taux régionaux (Aube : 72%, 10 textes *vs* moy. = 71%), tandis que dans la première moitié du XIII^e s., l'Aube de nouveau (100%, 2 textes), l'Oise (94%, 2 textes) et la Normandie (89%, 5 textes) ont un taux régional moyen supérieur à la moyenne des taux régionaux (moy. = 80%), et c'est encore le cas de l'Oise (78%, 2 textes *vs* moy. = 74%) après 1250. S'il y a donc quelques déplacements dans les frontières précises de cette 'vallée' de taux plus fiables, celles-ci sont concentrées dans la première moitié du XIII^e s. et n'invalident pas fondamentalement le dessin de la diagonale observée.

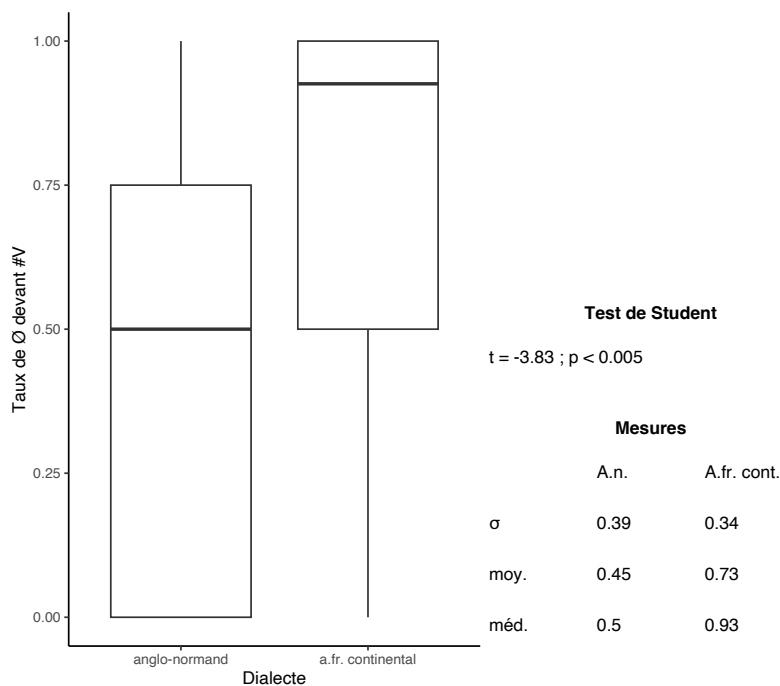


FIGURE 10.9 – Diachronie des taux de défaut de <e>, anglo-normand *vs* ancien français continental, conjonctions en « base + -que » devant initiale vocalique.

En revanche, si la diatopie semble peu sensible à la diachronie, l'inverse n'est pas vrai : les taux extrêmement bas (relativement à la moyenne des taux régionaux) de l'anglo-normand, observables en fig. 10.7 et 10.8, conjugués à la précocité du répertoire anglo-normand au sein du corpus, concourent à la tendance à la hausse observée en fig. 10.4.b,c. Ainsi, comme représenté en fig. 10.9, la distribution des taux anglo-normands et celle des taux continentaux est partiellement symétrique : la médiane des taux anglo-normands se situe à la frontière entre le premier et le second quartile des taux continentaux, tandis que la médiane des taux continentaux se situe bien au-dessus de la frontière entre le troisième et le quatrième quartile anglo-normand. Si l'effet, tel que mesuré par le test de Student, est faible ($t = -3,9$), il est néanmoins significatif ($p < 0,005$), et sa faiblesse provient ici de la distribution des deux jeux de données sur l'ensemble de l'échelle des ordonnées.

De fait, si l'on isole les données continentales tel qu'en fig. 10.10.b, la tendance générale n'est plus à la hausse mais à la baisse. Ce sont donc

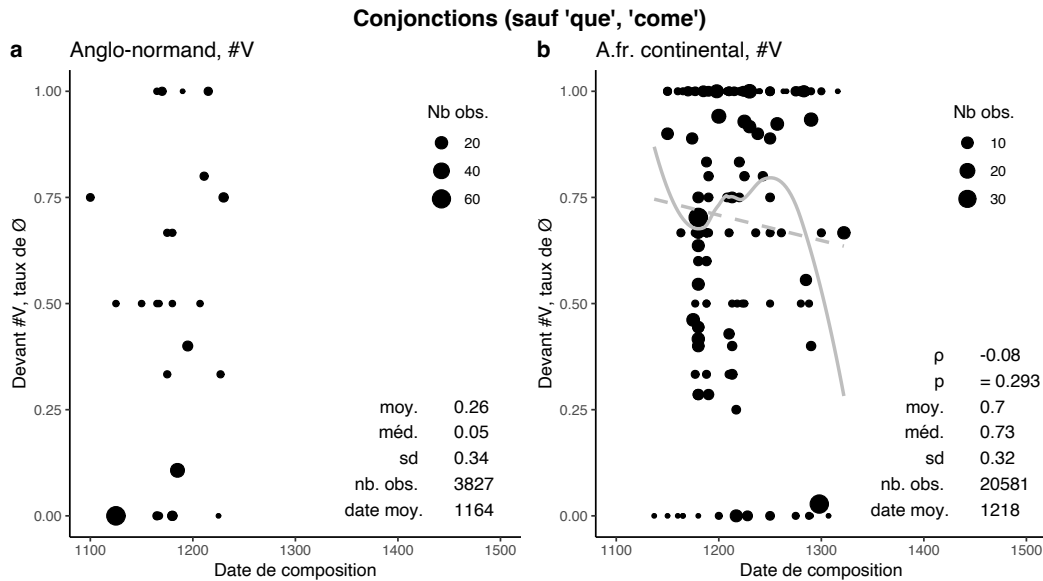


FIGURE 10.10 – Diachronie des défauts de <e>, anglo-normand *vs* ancien français continental, conjonctions en « base + -que » devant initiale vocalique.

les taux anglo-normands bas et précoces qui génèrent, à eux seuls, la tendance régulière à la hausse observée *supra*. Néanmoins, la conformité de la droite en fig. 10.10.b au modèle de régression linéaire est beaucoup trop faible ($\rho = -0,008$) et non significative ($p = 0,3$) pour qu'on puisse conclure à une tendance généralisée à la baisse des taux d'élision, mais ces mesures permettent d'établir qu'il ne convient donc pas de conclure que l'élision des conjonctions en *base + -que* est de plus en plus systématique.

10.2.2.4 En somme

Les conjonctions formées par l'adjonction de *-que* à une base sont sensibles à l'élision mais celle-ci n'est pas systématique, et sans que se dégage une évolution diachronique de cette évolution qui ne soit pas due à une interférence diatopique. En revanche, la distribution diatopique des occurrences élidées est remarquable, avec soit une diagonale Nord-Ouest - Sud-Est soit une large moitié Sud-Est du domaine dont les taux sont les plus faibles. Cette distribution est relativement stable en diachronie.

10.2.3 L'élément *que*

Cette section s'attache à décrire les alternances entre, d'une part, les types *qui* et *que*, et d'autre part le type *qu*. Elle traite successivement de la distribution diachronique et contextuelle des variantes du type *qu* (sec. 10.2.3.1), puis de leur distribution diatopique (sec. 10.2.3.2) et enfin de l'intersection entre les distributions diatopiques et diachroniques (sec. 10.2.3.3).

10.2.3.1 Distribution diachronique et contextuelle

La quantité absolue des occurrences de l'élément *que*, c'est-à-dire des types *qui*, *que* et *qu* annotés pos = "PROCON" est bien plus élevée que celle des locutions conjonctives en *base* + *-que* : on dénombre 30 fois plus d'occurrences pour l'élément *que* que pour les locutions. Comme représenté sur les graphiques de séries temporelles et sur les boîtes à moustache en fig. 10.11.a-e en page suivante, les occurrences du type *qu* (*k*, *q*) sont extrêmement marginales devant initiale consonantique (moy. < 0,01%), alors qu'elles sont bien plus représentées devant initiale vocalique, avec une moyenne de 41% par texte, et une moyenne pondérée par le nombre d'occurrences des textes de 39%. Néanmoins, si l'effet du contexte est significatif ($p < 0,005$), il est modéré ($t = 32$), étant donné que l'élision n'est systématique dans aucun texte et est absente dans plusieurs textes.

Ces éléments font directement signe vers la nature complexe de cet élément *que*, et vers le fait que nos données pour cet élément mélangent pronoms et conjonctions pures, et cas sujet et cas régime pour le pronom.²⁹ Comme mentionné en sec. 10.1.2 *supra*, dans la mesure où les occurrences élidées sont toujours annotées avec l'un des lemmes du type *que*, il est impossible de ré-introduire ces distinctions dans nos données. De fait, ces données ne représentent qu'un *proxy* pour l'étude de la conjonction *que*, et nous ne pouvons exclure que la différence de taux entre l'élément *que* et le *-que* des locutions conjonctives ne soit due qu'à cette ambiguïté morphosyntaxique de l'élément *que*.

29: Les mots dont la partie du discours est PROCON ne sont pas annotés en cas dans le corpus, contrairement aux mots annotés comme des pronoms.

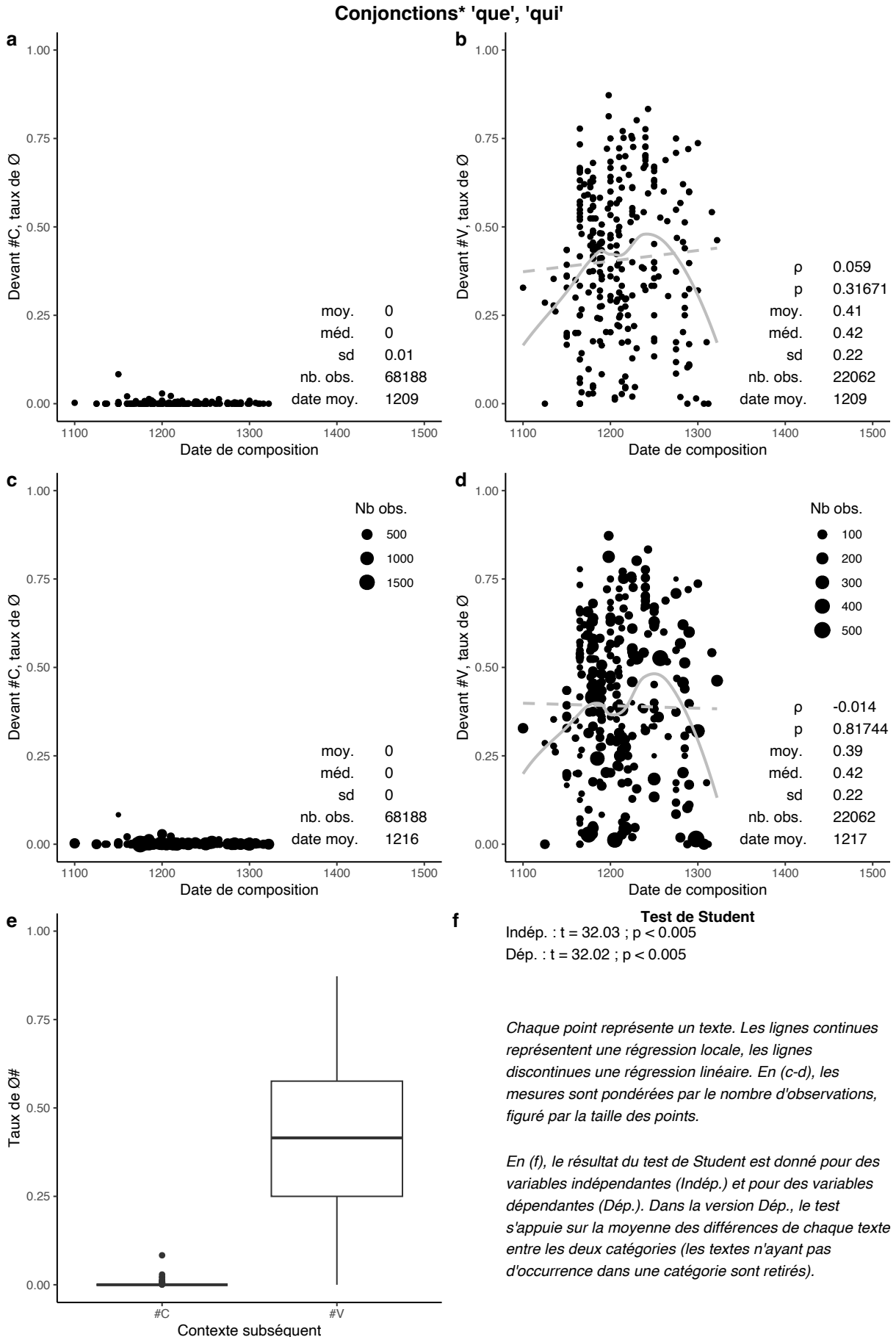


FIGURE 10.11 – Distribution contextuelle des variantes de conjonctions* *que* en défaut de <V>.
 *Conjonction ou autre nature : élément *que* (dont *qui*).

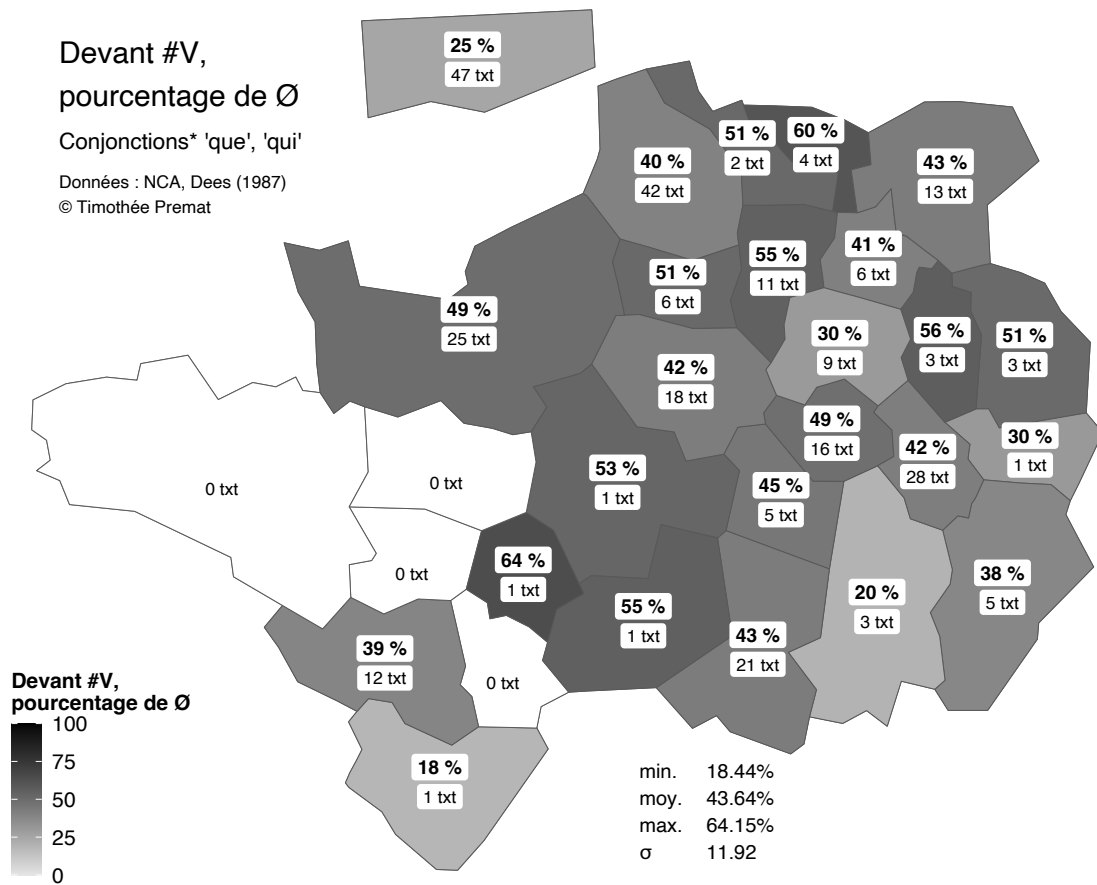


FIGURE 10.12 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, pour l'élément *que*, suivi d'une initiale vocalique.

*Occurrences annotées pos = "PROCON" des types *que*, *qui* et *qu*.

10.2.3.2 Distribution diatopique

La carte en fig. 10.12 représente la distribution diatopique des élisions de l'élément *que*. Sur cette carte, la situation est moins hétérogène ($\sigma = 12$) qu'elle ne l'est pour les locutions en *base + -que* ($\sigma = 19$, en fig. 10.5 *supra*). Si l'on retrouve sur cette carte certaines régions appartenant, pour les conjonctions en *base + -que*, à la diagonale à taux inférieurs à la moyenne, cette correspondance n'est pas systématique. Ainsi, l'Angleterre (25%), la région parisienne (42%), la Marne (30%), la Haute-Marne (42%) et la Bourgogne (20%) pratiquent moins l'élision que la moyenne des régions (moy. = 44%). Ce n'est en revanche plus le cas de l'Aube (49%) ou de la Normandie (49%), et l'on retrouve hors de la diagonale Nord-Ouest - Sud-Est peuvent avoir des taux inférieurs à la moyenne, comme c'est le cas de la Charente-Maritime (18%), des Vosges (30%), de la région Somme, Pas-de-Calais (40%) ou des Ardennes (41%).

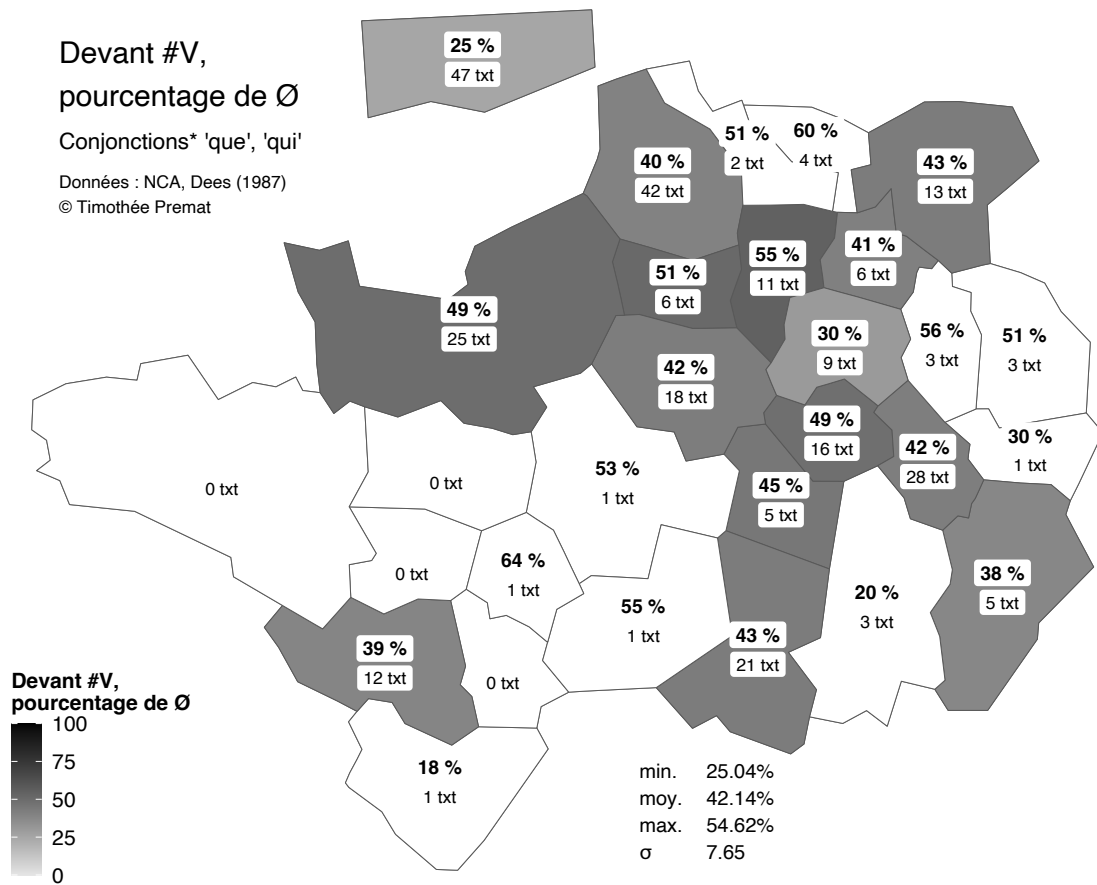


FIGURE 10.13 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, élément *que* suivi d'une initiale vocalique et à l'omission des régions disposant de moins de cinq textes.

*Occurrences annotées pos = "PROCON" des types *que*, *qui* et *qu*.

S'il y a donc une forme de continuité entre la distribution diatopique de l'élision des conjonctions en *base* + *-que* et de l'élément *que*, cette continuité n'est que partielle. Si l'on restreint la carte aux régions disposant de 5 textes ou plus, comme représenté en fig. 10.13, la continuité s'efface, les régions ayant plus de quatre textes formant un ensemble plus homogène ($\sigma = 8$) sur lequel la diagonale n'est plus isolée. Si nous ne pouvons exclure catégoriquement le fait que l'augmentation du nombre de textes soit responsable de ce nivellement des taux, il faut néanmoins remarquer que ce n'était pas le cas pour les conjonctions en *base* + *-que* en fig. 10.6 *supra*, où la restriction aux régions disposant de 5 textes ou plus conservait des valeurs contrastées et maintenait au moins partiellement le dessin de la diagonale Nord-Ouest - Sud-Est, aussi cette hypothèse n'est pas la plus probable. Il semble donc préférable d'admettre que la variation de l'élément *que*, si elle dispose de caractéristiques communes avec la variation des conjonctions en *base* + *-que*, diffère de celle-ci sur d'autres points, ce qui rend la 'vallée' diagonale discontinue et moins lisible.

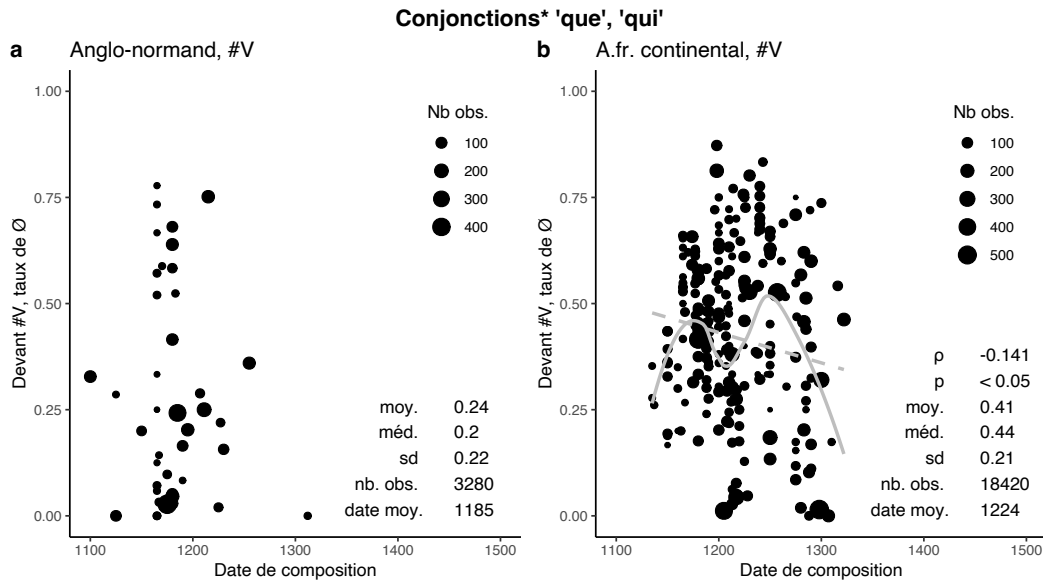


FIGURE 10.14 – Diachronie des défauts de <e>, anglo-normand *vs* ancien français continental, élément *que*, suivi d’une initiale vocalique.
* Occurrences annotées pos = "PROCON" des types *que*, *qui* et *qu*.

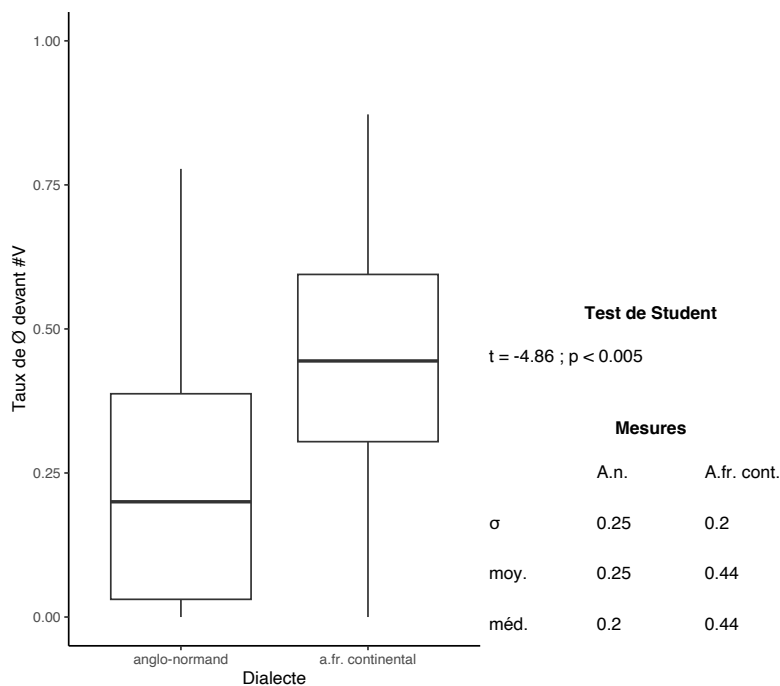


FIGURE 10.15 – Diachronie des taux de défaut de <e>, anglo-normand *vs* ancien français continental, élément *que*, suivi d’une initiale vocalique.

Enfin, il faut noter que l’Angleterre est la seule région dont le taux est inférieur à 20% à disposer d’un nombre conséquent de textes. Cette région dispose pourtant de 11 textes (sur 47) ayant plus de 50% d’élisions, comme représenté en fig. 10.14. Néanmoins, son taux moyen résulte d’une concentration élevée de textes dont le taux est bas, la médiane anglo-normande s’établissant à 20% (*vs* 44% sur le continent).

Il en résulte, comme représenté en fig. 10.14, qu'à peu près les deux tiers des textes anglo-normands se situent dans la zone de taux bas qui correspond au premier quartile des textes continentaux. De fait, si l'anglo-normand n'est pas ici significativement plus ou moins homogène que le domaine continental (a.n. : $\sigma = 0,25$; a.fr.cont. : $\sigma = 0,2$), son centre de gravité se situe dans des taux bien plus bas que le domaine continental.

10.2.3.3 Intersection de la diatopie et de la diachronie

L'hétérogénéité constatée sur la carte synchronique en 10.12 se retrouve en partie en diachronie, comme le montrent les cartes en fig. 10.16 et 10.17 en pages suivantes. Ainsi, alors que le Nord et le Nord-Est du domaine ne pratiquent que faiblement l'élision au début du corpus (fig. 10.16.a : Meuse et région Somme, Pas-de-Calais : 19% ; Angleterre : 20% ; Wallonie : 26% ; Meuse : 28%), les taux de ces régions augmentent dès la seconde moitié du XII^e s., à l'exception de celui de l'Angleterre (fig. 10.16.b : Meuse : 75% ; région Somme, Pas-de-Calais : 47% ; Angleterre : 25% ; Wallonie : 42% ; Meuse : 75%). Hormis cette caractéristique de rapprochement ou de dépassement de la moyenne du Nord-Est au cours du XII^e s., l'examen des taux moyens de chaque région à chaque période conclut surtout à une grande instabilité. Le mitage territorial constaté sur la carte synchronique est donc également valable en diachronie, représentant une variation en apparence assez libre.

Enfin, dans le cas de l'Angleterre, et contrairement à ce qui a pu être avancé pour d'autres parties du discours, l'intersection entre diachronie et diatopie est faible. Certes, les taux anglo-normands sont faibles et majoritairement précoces, mais leur faiblesse se maintient au XIII^e s. (fig. 10.17) et, comme illustré sur le graphique de séries temporelles en fig. 10.14 *supra*, l'anglo-normand dispose de taux inférieurs à ~20% dans la seconde moitié du XII^e s, ce qui n'est pas le cas sur le continent. La spécificité anglo-normande est donc peu sensible à la diachronie de ce point de vue.

10.2.3.4 En somme

L'élément *que*, comme les conjonctions en *base + -que*, est sensible à l'élision : les défauts vocaliques ne se produisent que devant initiale vocalique. Le taux d'élision est néanmoins plus faible pour l'élément *que* (41%) qu'il ne l'est pour les locutions en *base + -que* (69%), et il est possible que cela provienne de notre incapacité à dissocier conjonctions pures et formes pronominales. La distribution diatopique des élisions de l'élément *que* est bien moins homogène que celle des locutions en *base + -que*, et la 'vallée' diagonale observée pour les locutions n'apparaît plus ici que sous une forme discontinue, et contrastant moins fortement avec ce qui l'entoure. Enfin, cette hétérogénéité spatiale vaut également en diachronie, et ce dans deux dimensions : non seulement chaque demi-siècle est relativement hétérogène, mais il n'y a pas non plus, hormis dans le cas de l'Angleterre, de continuité forte entre les différentes périodes.

Si l'on admet qu'une partie non négligeable des occurrences non élidées provient de ce que l'occurrence assure un rôle pronominal qui dépasse

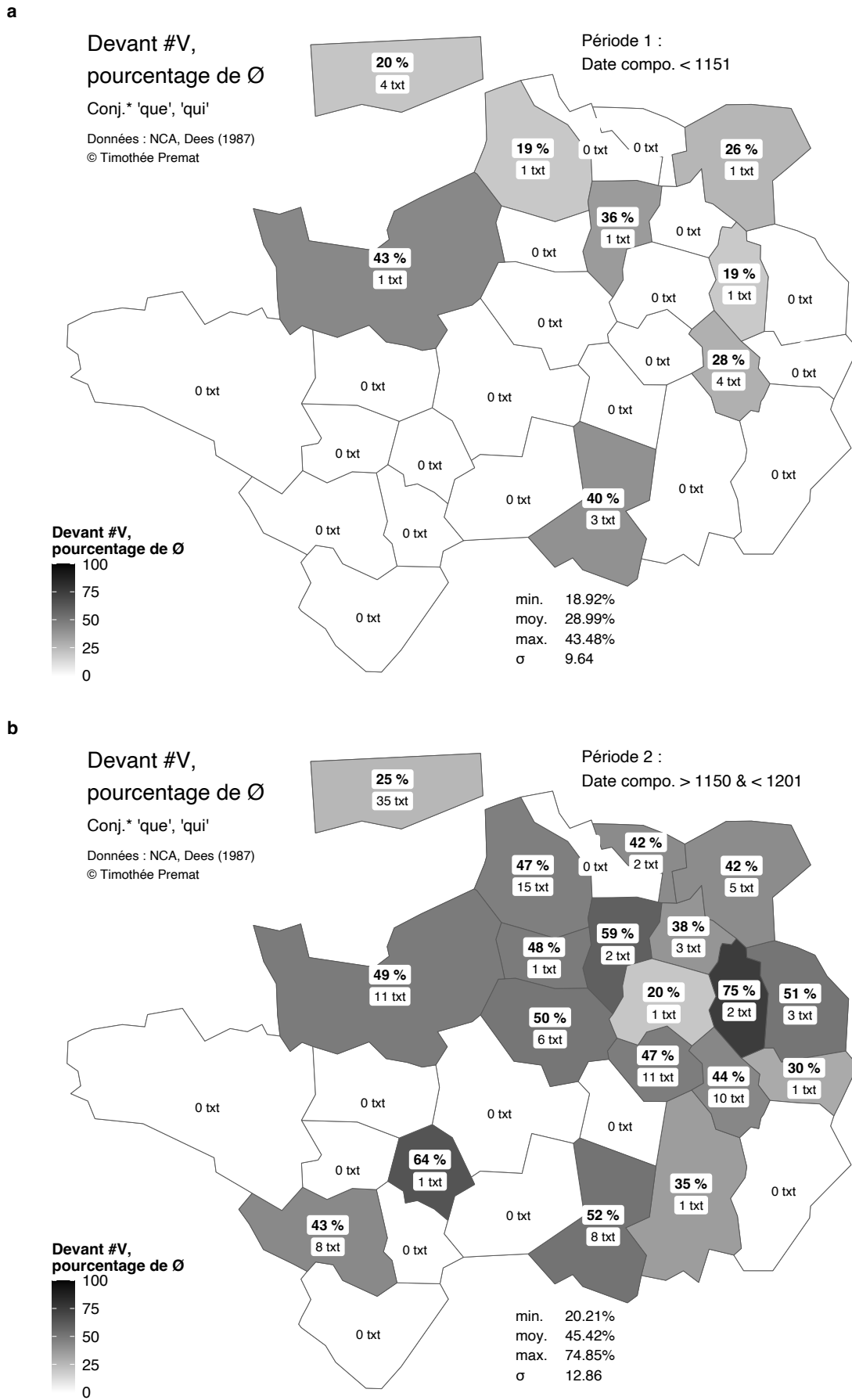
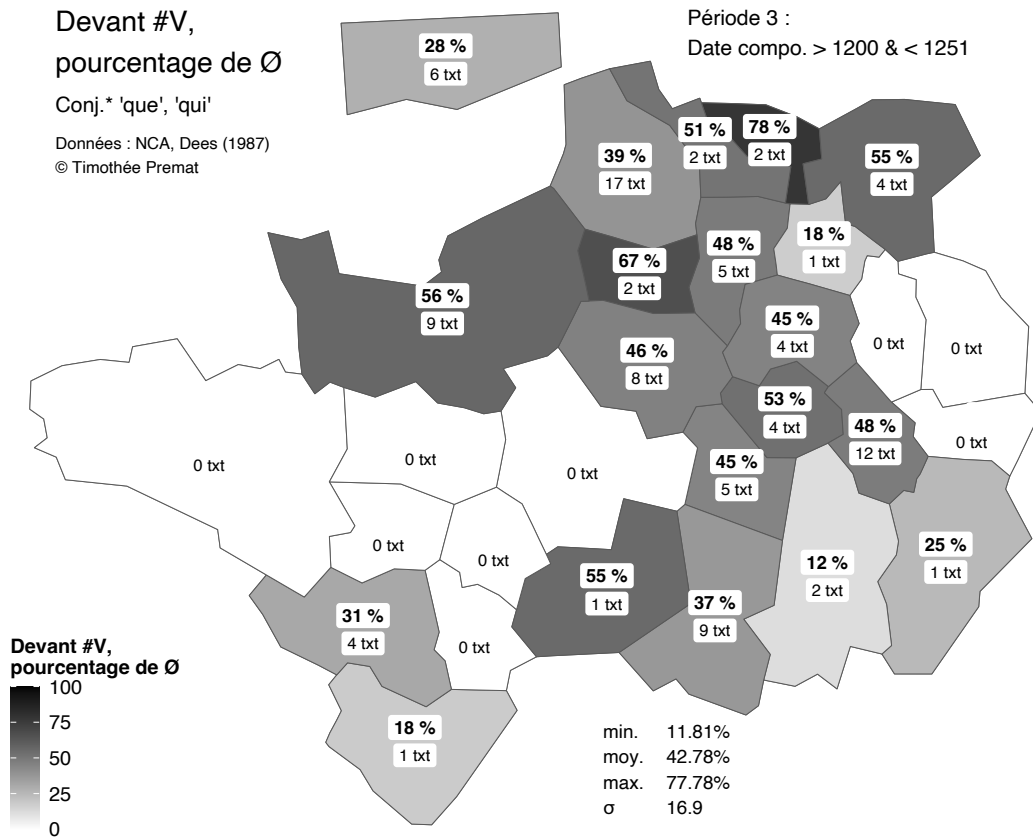


FIGURE 10.16 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1150) et 2 (b : 1151-1200), élément *que* suivi d’une initiale vocalique.

*Occurrences annotées pos = "PROCON" des types *que*, *qui* et *qu*.

a



b

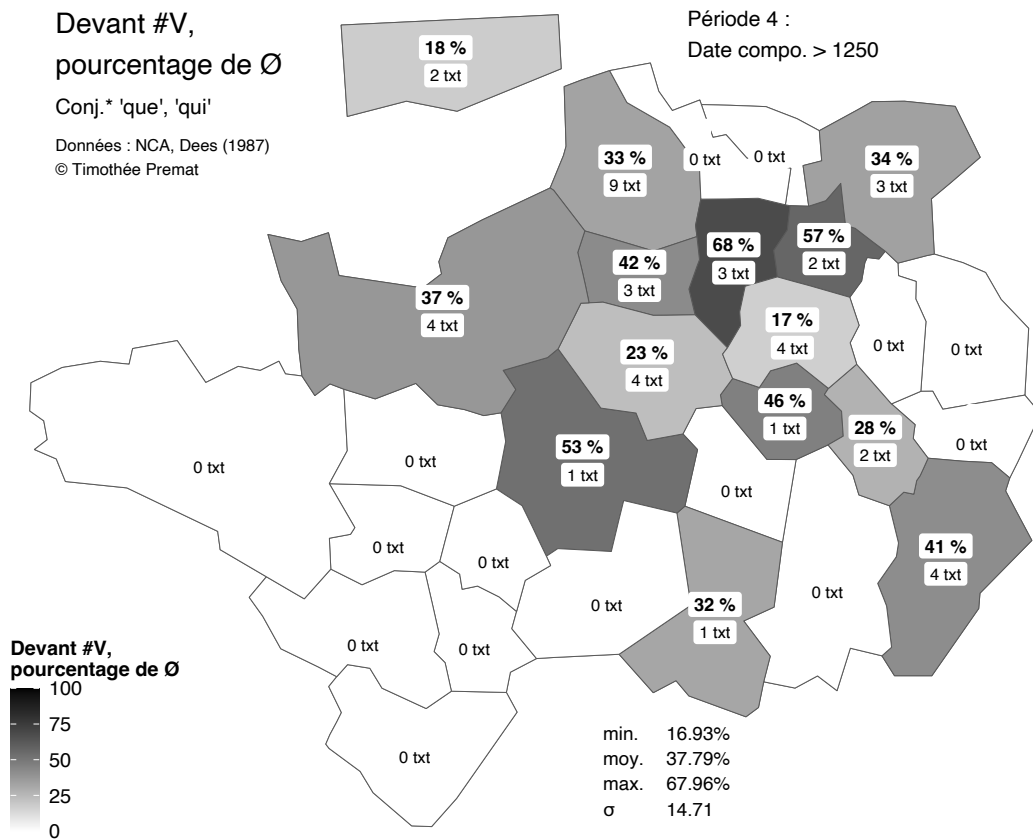


FIGURE 10.17 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), élément *que* suivi d’une initiale vocalique.

*Occurrences annotées pos = "PROCON" des types *que*, *qui* et *qu*.

son statut simplement conjonctif, la distinction entre le traitement du *que* des locutions en *base + -que* et du *que* autonome de l'élément *que* peut être réduite à un simple critère de biais dans les données. Dans ce cas, c'est bien le *-que* des locutions qui représenterait le comportement du *que* purement conjonctif, tandis que les données de l'élément *que* seraient inconclusives pour ce qui est du statut de la conjonction. Cette hypothèse permettrait de rendre compte du fait qu'aucun texte ne pratique l'élision systématique de l'élément *que*, ce qui est remarquable.

Néanmoins, cette hypothèse doit intégrer le fait que dans près des trois quarts des textes, le *-que* des locutions n'est pas systématiquement élide non plus. Cette hypothèse n'est, au demeurant, pas testable en l'état du corpus et de nos moyens informatiques et, au vu du nombre d'occurrences (plus de 22 000 occurrences devant initiale vocalique), impossible à vérifier par une inspection manuelle des contextes. Il convient donc de constater la différence des taux d'élision entre l'élément *que* et le *-que* des locutions conjonctives, et de considérer que cette différence est peut-être due au statut pronominal de certaines occurrences de l'élément *que*, mais sans que nous puissions le démontrer ou l'infirmier.

10.2.4 *Com(e)*

Cette section traite des alternances entre les formes du type *come* et celles du type *com*. Comme expliqué en sec. 10.1.3, cette alternance ne tient pas à ce que nous avons défini comme étant un défaut de <e>, dans la mesure où l'étymologie et les lois de la phonétique historique ne génèrent, à partir de l'étymon populaire *como*, en principe que les formes monosyllabiques du type *com*. Néanmoins, les formes bisyllabiques sont attestées dès *Roland*, d'abord en distribution syntaxique complémentaire avec les formes monosyllabiques, avant que cette distribution complémentaire ne s'affaiblisse et se neutralise, alors que les formes bisyllabiques remplacent progressivement les formes monosyllabiques. La première section (sec. 10.2.4.1) interroge la distribution des formes avec et sans <e> final dans des contextes où seules des formes sans <e> sont attendues selon la règle archaïque, tandis que la seconde section (sec. 10.2.4.2) permet la comparaison avec la variation sans considération pour le contexte syntaxique dans lequel elles apparaissent.

10.2.4.1 Alternance *com* ~ *come* en contexte syntaxique de *com*

Tous les graphiques et toutes les mesures de cette section concernent, comme ailleurs dans cette thèse, les taux de formes sans <e>. Comme, ici, il s'agit de mesurer l'immixtion de formes du type *come* au sein de contextes syntaxiques où seules les formes du type *com* devraient se trouver, cette immixtion se mesure comme étant l'opposée des taux d'absence de <e> mesurés par Comparalem : plus un taux de défaut de <e> est bas, plus le type *come* s'immisce dans le contexte syntaxique du type *com*. Cette section expose d'abord l'évolution diachronique de la neutralisation de l'opposition syntaxique entre les deux types, en prenant d'emblée en compte l'influence du contexte phonologique, avant de détailler la diatopie de cette neutralisation.

Distribution diachronique et contextuelle de la neutralisation. En sec. 10.14, nous avons illustré l'affaiblissement de la distribution syntaxique complémentaire entre le type *com* et le type *come* en documentant la contamination par le type *come* de deux contextes en principe dévolus au type *com*, qui sont la précession d'un pronom personnel sujet ou d'un verbe conjugué. Cette illustration avait été produite, en fig. 10.3, sans distinction en fonction du contexte phonologique. La fig. 10.18, en page suivante, reprend le même principe, mais avec cette fois-ci une sensibilité au contexte phonologique. Elle présente le taux de formes du type *com* devant pronom personnel sujet et verbe conjugué, mais en séparant les occurrences selon que l'initiale du pronom ou du verbe est vocalique ou consonantique.

La dissociation des contextes provoque une accentuation artificielle de la pente de la droite de régression, mais les données sont de toute manière peu conformes au modèle de la régression linéaire, le coefficient de corrélation était trop faible ($\rho = -0,2$ __#C; $\rho = -0,3$ __#V). On note en revanche, dès le début de la seconde moitié du XII^e s., que de nombreux textes n'ont plus que le type bisyllabique dans ces contextes syntaxiques en principe réservés au type monosyllabique, et ce dans les deux contextes phonologiques. Enfin, il importe de noter que le type monosyllabique se fait plus rare, dans ces contextes syntaxiques qui lui sont réservés, devant initiale vocalique (68% du type *com*) que devant initiale consonantique (82%); l'effet mesuré par le test de Student est certes peu marqué ($t = -4$) mais statistiquement significatif ($p < 0,005$).

De fait, dans les deux contextes phonologiques, plus de la moitié des textes ne disposent que d'occurrences monosyllabiques, ne contredisant pas la distribution complémentaire de Roland (méd. = 100%). Néanmoins, on note sur le graphique en boîtes à moustache en fig. 10.18.e que si les trois quarts des textes ont plus de 75% du type *com* dans ces contextes syntaxiques (seul le premier quartile s'étend en-deçà de 75%), ce n'est pas le cas devant initiale vocalique, où le second quartile s'étend de ~25% à 100%. Ainsi, la différence entre les taux moyens devant initiale consonantique et devant initiale vocalique tient au fait que, devant initiale vocalique, une plus grande quantité de textes ignore la distribution syntaxique; mais dans les deux cas la plupart des textes ont bien le type *com* devant pronom personnel sujet ou verbe conjugué.

Ce point peut sembler contre-intuitif, dans la mesure où c'est devant initiale consonantique qu'on trouve le plus de *com*. Mais cette distribution prend sens si l'on considère que la majorité du corpus du NCA est versifiée, et que les occurrences notées par VISING (1895) dans Roland montrent que l'alternance *cum* ~ *cume* n'est pas une licence poétique mais peut au contraire briser l'isosyllabisme d'un vers. De fait, c'est uniquement devant consonne qu'un vers peut devenir hypermétrique par l'emploi d'un *come* à la place d'un *com* (en tête d'une proposition dont le verbe est exprimé). Aussi, dans le processus de copie d'un texte, si l'alternance contextuelle s'est affaiblie dans la grammaire d'un scribe donné, celui-ci est moins susceptible d'écrire *come* devant consonne que devant voyelle, dans la mesure où remplacer un *com* par un *come* n'a d'incidence métrique que devant initiale consonantique. C'est donc dans le contexte où la phonologie est susceptible d'effacer la différence entre le type monosyllabique et le type bisyllabique (/kom/ et /komə/ prononcés

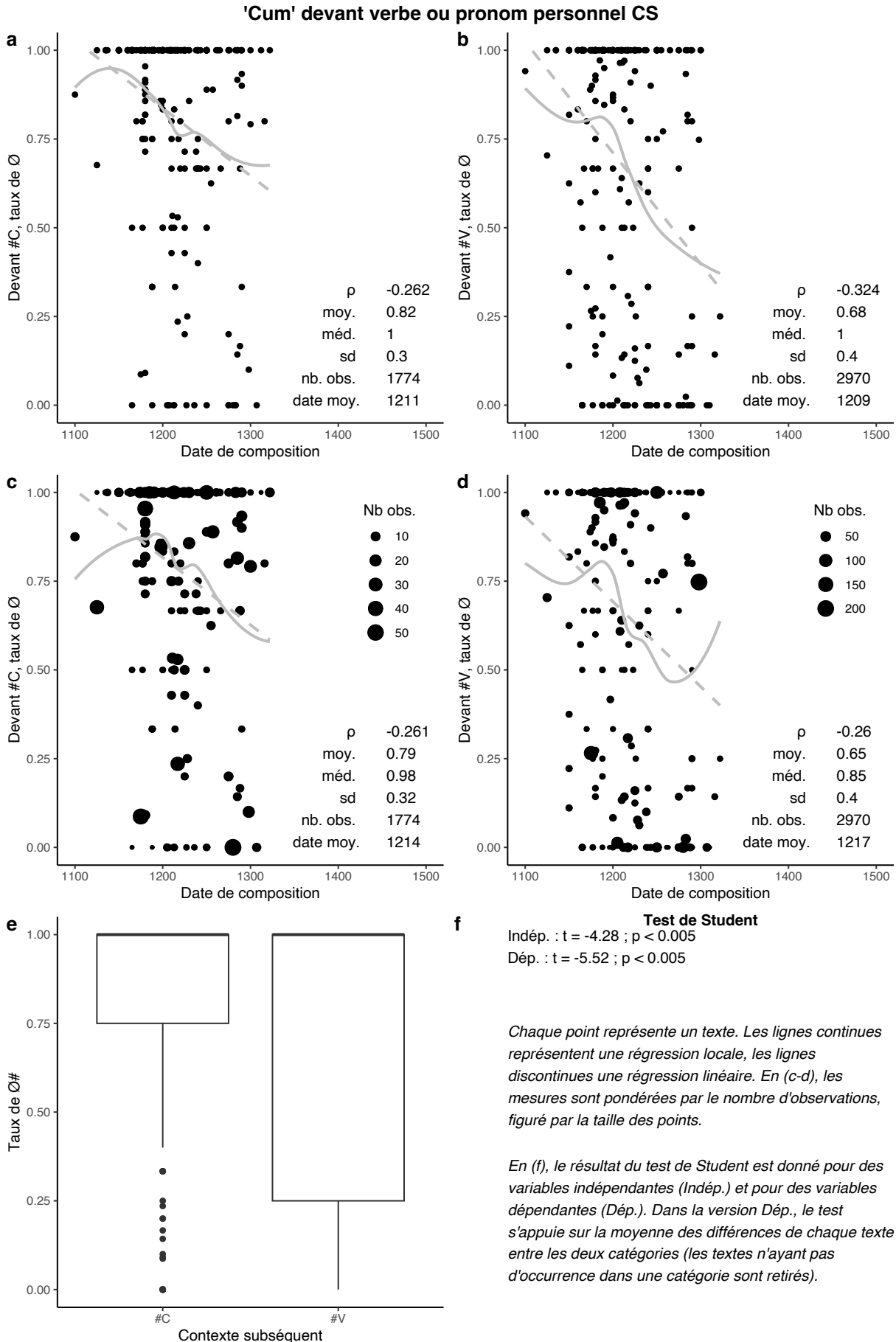


FIGURE 10.18 – Distribution contextuelle des variantes de conjonctions « come » en défaut de <e> devant pronom personnel sujet ou verbe conjugué.

[kom] devant initiale vocalique) que l'on voit l'affaiblissement le plus marqué de la distribution syntaxique.

— *Début de la neutralisation.* En diachronie, les deux premiers textes à disposer de formes bisyllabiques dans ce contexte précis sont *Roland* et *psautier*, ce dernier contenant 11 occurrences *cume* suivies d'un pronom personnel sujet pour 29 occurrences *cum* dans ce même contexte. Sur ces onze occurrences *cume*, 6 sont devant initiale vocalique (10.22), et cinq sont devant initiale consonantique (10.23) ce qui, rapporté au nombre total d'occurrences du lemme dans ces contextes, ne favorise ni l'un ni l'autre des deux contextes.

- (10.22) *cume eles seient lances jete sur le seignur*
 « comme si elles [= les paroles] étaient des lances jetées sur le seigneur »
 (psautier, p. 45, trad. pers.)
- (10.23) *cume jeo crierai a tei cume je leverai mes mains*
 « comme je crierai ton nom, comme je lèverai les mains »
 (psautier, p. 20, trad. pers.)

Ce texte, daté dans le NCA de 1125 avec une marge d'incertitude de 25 ans, est peut-être légèrement postérieur au *Comput* de Philippe de Thaon (que la BFM date de 1113 ou 1119),³⁰ que SUCHIER (1879) et VISING (1895) citent comme n'appliquant plus de distinction entre *cum* et *cume*. Le témoignage de *psautier* est donc plus modéré que celui du *Comput* : il témoigne d'un certain affaiblissement de l'opposition syntaxique entre *cum* et *cume* en anglo-normand, mais cet affaiblissement ne rend pas encore les deux formes interchangeable : 72,5% des occurrences dans ce contexte demeurent *cum*, devant initiale vocalique comme devant initiale consonantique.

Les premiers textes disposant de 100% de formes bisyllabiques devant pronom personnel sujet ou verbe conjugué apparaissent au milieu du XII^e s. Entre 1150 et 1175, 51 textes contiennent des occurrences de « come » dans ces contextes syntaxiques, pour un total de 587 occurrences. L'augmentation de la quantité de textes permet donc, dès le troisième quart du XII^e s., une étude quantitative plus fournie. Sur ces 51 textes, 21 contiennent des occurrences du type *come* et 48 contiennent des occurrences du type *com*. Parmi les 21 textes contenant des occurrences bisyllabiques, seulement trois ne contiennent aucune occurrence monosyllabique (*deusamS*, *guigS*, *lanvalS*), tandis que parmi les 48 textes comportant des occurrences monosyllabiques, 30 ne disposent d'aucune occurrence bisyllabique. On voit donc apparaître la neutralisation de l'opposition *com* ~ *come*, non plus seulement sous la forme d'une marge d'erreur qui serait de plus en plus grande, mais sous la forme d'une domination du type *come* dans ce contexte syntaxique dans trois textes, accompagnée par une variation entre les deux formes dans 18 textes, tandis que 30 textes ne connaissent encore que le type monosyllabique devant pronom personnel sujet ou verbe conjugué.

psautier : ANON. (1125*). *Livre des psaumes*. Dees nil; compo. a.n., ca 1125; ms. a.n., ca 1150; CR2 (MICHEL 1876), DEAF : PsCambrM.

30: <http://catalog.bfm-corpus.org/compu>

deusamS : MARIE DE FRANCE (1165*). *Les Deus Amanz*. Ms. S; Dees : nil; compo. a.n., ca 1165; ms. frc., ca 1300; MS (transcr. inédite sans transcripteur indiqué).
guigS : MARIE DE FRANCE (1165*). *Le Lai de Guigemar*. Ms. S; Dees : nil; compo. N.-O., ca 1165; ms. frc., ca 1300; nil (transcr. inédite sans transcripteur indiqué).
lanvalS : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. S; Dees : nil; compo. N.-O., ca 1165; ms. frc., ca 1300; MS2 (MARIE DE FRANCE 1958); DEAF : MarieLanvR².

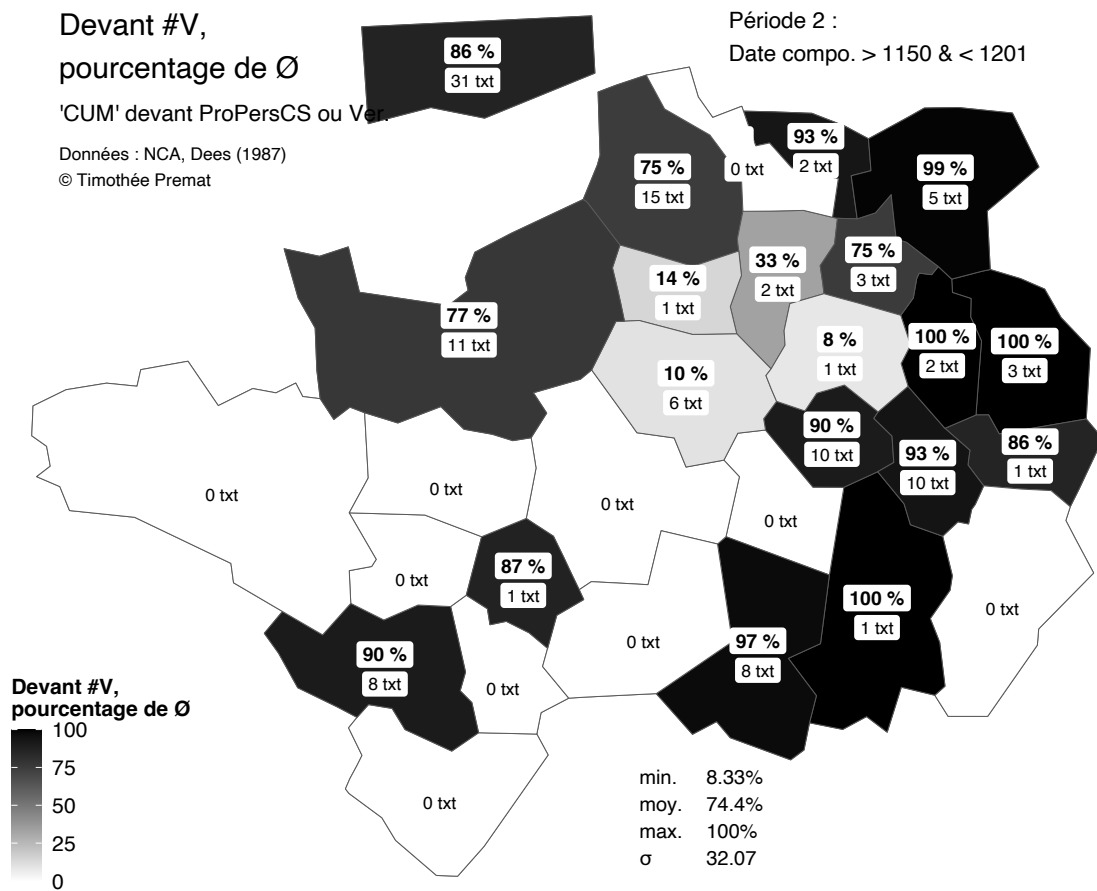


FIGURE 10.19 – Distribution diatopique des alternances <e>/Ø, période 2 (1151-1200), conjonction « come » devant pronom personnel sujet et verbe conjugué à initiale vocalique.

Distribution diatopique de la neutralisation. La primauté de la disparition de la distribution complémentaire syntaxique n'est pas réservée à l'anglo-normand ; sur les trois premiers textes n'ayant que des formes bisyllabiques dans ce contexte, l'un est anglo-normand (*deusamS*), tandis que les deux autres (*guigS* et *lanvalS*) sont donnés par la bibliographie philologique comme ayant été composés dans le Nord-Ouest et copiés en francien, c'est-à-dire sans trait dialectal marqué. C'est plutôt au centre du domaine que les occurrences bisyllabiques sont, proportionnellement, les plus nombreuses devant pronom personnel sujet ou verbe conjugué, ainsi que représenté par les cartes en fig. 10.19 (__#V) et 10.20 (__#C, page suivante), pour l'ensemble de la seconde moitié du XII^e s. Cet effet est plus marqué devant initiale vocalique (fig. 10.19), avec un pôle assez franc de textes où le type bisyllabique domine dans la Marne (8% de *com*), la région parisienne (10%), l'Oise (14%) et l'Aisne (33%).

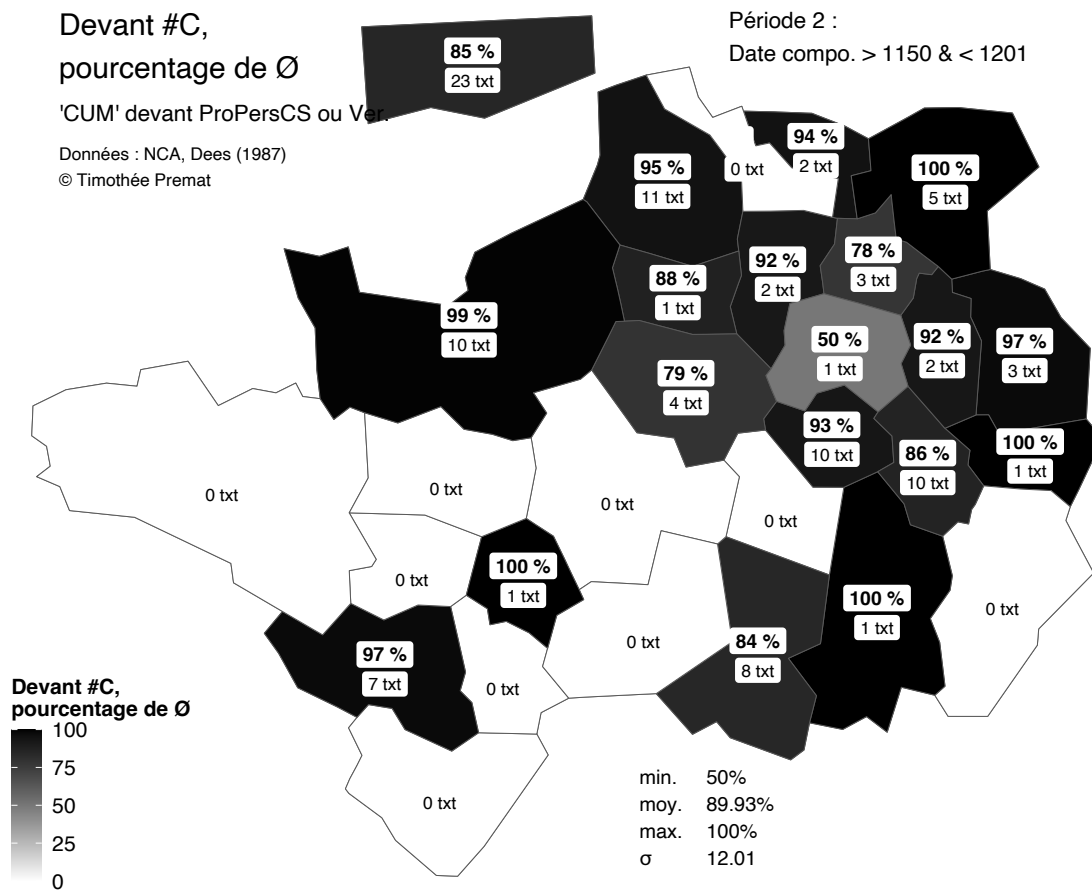
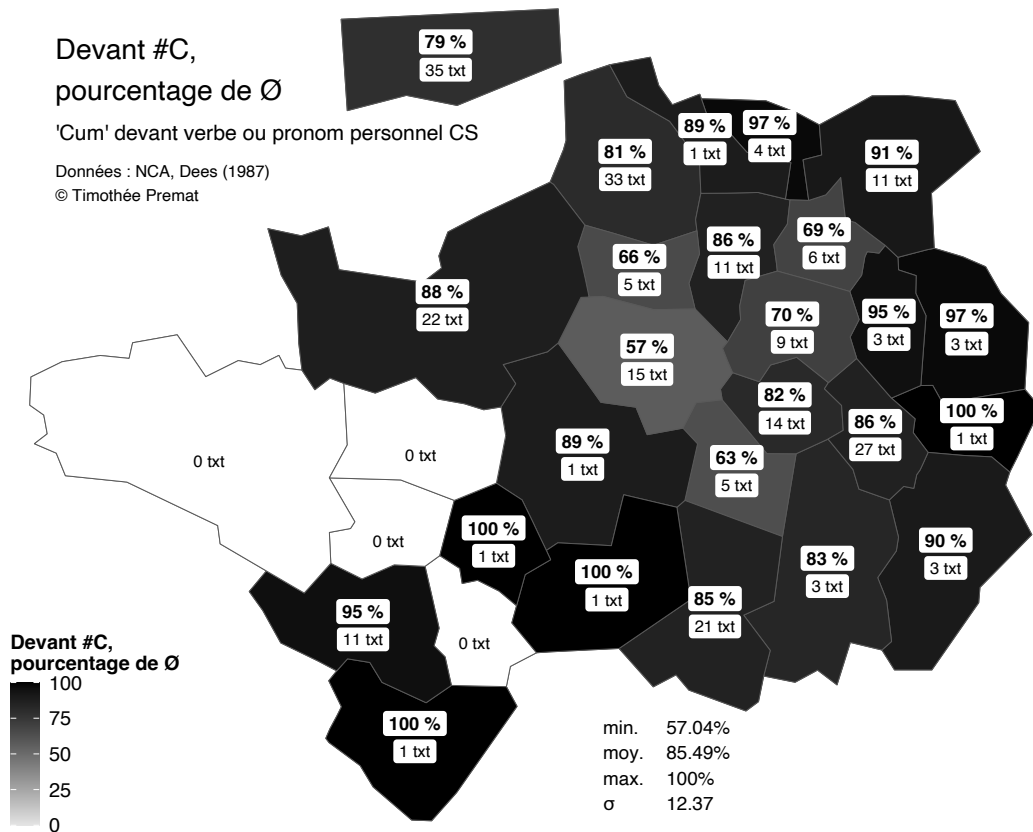


FIGURE 10.20 – Distribution diatopique des alternances <e>/Ø, période 2 (1151-1200), conjonction « come » devant pronom personnel sujet et verbe conjugué à initiale consonantique.

Devant initiale consonantique (fig. 10.20), seule la Marne se distingue clairement (50%, sur un seul texte), même si l'on note que la région parisienne (79%) demeure également en-deçà de la moyenne des taux régionaux (moy. = 90%). La remarque de VISING (1895, p. 120-123), selon laquelle une bonne conservation de la distinction syntaxique est observée à l'Est et au Nord mais pas en anglo-normand, ne doit donc pas être lue comme un effondrement de la distinction en anglo-normand plus vite que sur le continent : c'est bien au centre du domaine que l'on trouve le plus d'occurrences du type *come* dans ces deux contextes réservés *a priori* à *com*, tandis que l'Angleterre conserve un taux moyen proche de celui de *Roland* dans la seconde moitié du XII^e s.

Cette distribution diatopique, relativement stable dans le temps, se retrouve également en synchronie, comme le montrent les cartes en fig. 10.21 page suivante. La distribution syntaxique, qui évite la forme bisyllabique en introduction d'une proposition dont le verbe est exprimé, semble donc se maintenir, en moyenne, aux marges du domaine, tandis qu'elle s'effondre au centre de celui-ci. L'extension du centre et des marges dépend du contexte, avec une faible taille de la zone centrale où

a



b

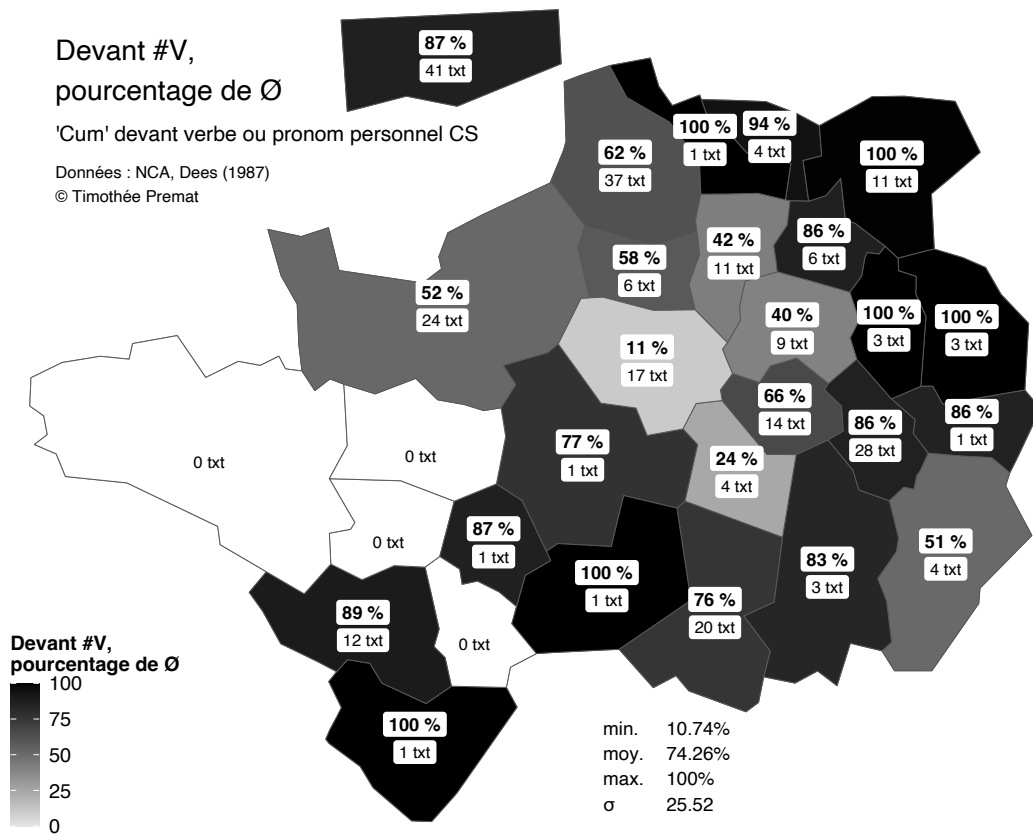


FIGURE 10.21 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, conjonctions « come » suivies d'une initiale consonnantique (a) et suivies d'une initiale vocalique (b).

la distribution syntaxique est neutralisée devant initiale consonantique (fig. 10.21.a), et au contraire un large centre et des marges réduites devant initiale vocalique (fig. 10.21.b). Dans les deux contextes, l'émanation de la domination du type bisyllabique depuis le centre du domaine ne semble pas uniquement radiale mais également orientée selon un axe Nord-Ouest - Sud-Est, ce qui rapproche la diatopie de ces formes de l'axe observé clairement pour l'élision des conjonctions en *base* + *-que* (fig. 10.5 *supra*) et observable, certes bien moins clairement, pour l'élément *que* (fig. 10.12 *supra*).

Il ne semble donc pas qu'il faille conclure à une disparition homogène de la distribution syntaxique entre *com* et *come*. La neutralisation de leur opposition est certes attestée dès le *xii^e* s., et dans les deux contextes, mais la majorité des textes conserve une préférence, catégorique ou tendancielle, pour introduire les propositions dont le verbe est exprimé avec les formes monosyllabiques du type *com*. Il est néanmoins notable que cette majorité de textes se trouve rejetée aux marges du domaine, avec une extension de ces marges plus importante devant initiale consonantique que devant initiale vocalique.

En somme. Il existe dans les premiers textes une distribution complémentaire syntaxique qui restreint l'emploi du type bisyllabique *come* à l'introduction des seules comparaisons elliptiques; nous avons pu tester le degré de complémentarité de cette distribution devant pronom personnel sujet et verbe conjugué, qui représentent deux contextes non elliptiques, où c'est donc *com* qui est attendu. Cette distribution est respectée de manière catégorique dans certains textes, ou prend la forme d'une simple tendance dans d'autres, ce qui est le cas de *Roland*. Si certains textes n'emploient que le type monosyllabique *com* pour introduire une proposition non elliptique, et ce sur l'ensemble de la période, on note simultanément une tendance générale à la baisse de l'emploi du type *com*, au profit du type *come*, et l'existence de textes où seul *come* introduit des propositions dont le verbe est exprimé. De manière surprenante, c'est devant initiale vocalique que le type *come* remplace le plus le type *com*. Devant initiale consonantique, ce remplacement est modéré et surtout actif dans le centre du domaine, tandis que devant initiale vocalique, ce remplacement peut être bien plus fort et seules les marges conservent de manière plus ou moins catégorique l'emploi de *com* devant pronom personnel sujet ou verbe conjugué.

Il reste désormais à comparer ces données avec celles de l'ensemble des occurrences *com* et *come*, indépendamment de la partie du discours du mot subséquent. Cette comparaison permet de ne pas se limiter à l'évolution de la contamination des contextes de *com* par *come*, pour interroger les formes de l'ensemble du corpus. Elle permet également de ré-interroger les données restreintes à la précession d'un pronom personnel sujet ou d'un verbe conjugué, dans la mesure où celles-ci ne mesurent pas la co-existence générale des deux types dans le texte; or, si un texte ne connaît que des formes *cum*, la seule présence de *cum* en introduction d'une proposition dont le verbe est exprimé ne témoigne pas nécessairement d'un maintien de la distribution syntaxique.

10.2.4.2 Alternances *com* ~ *come* dans tous les contextes

Cette section étudie la distribution de l'ensemble des conjonctions « come », indépendamment de la partie du discours du mot suivant. Comme pour « come » devant pronom personnel sujet ou verbe conjugué, cette section expose la distribution diachronique et contextuelle des deux types, avant de traiter de leur distribution diatopique et, cette fois-ci, de l'intersection entre diatopie et diachronie.

Distribution diachronique et contextuelle Les graphiques de séries temporelles en fig. 10.4.a-d, page suivante, représentent le taux de défaut de <e> de la conjonction *come*, pour chaque texte, avec (fig. 10.4.c,d) et sans (fig. 10.4.a,b) pondération des taux par le nombre d'occurrences de chaque texte. Les données, ici, concernent l'ensemble des occurrences du lemme conjonctif, sans distinction en fonction de la partie du discours du mot suivant. De ce fait, dans l'analyse contextuelle, nous avons dû ré-introduire l'exclusion de <i> initial, ne pouvant contrôler systématiquement que ce <i> représente bien une voyelle et non une consonne. Vis-à-vis des données de la section précédente, où « come » précède une proposition dont le verbe est exprimé, on retrouve la même tendance générale à la baisse des formes monosyllabiques et à la hausse des formes bisyllabiques (cependant toujours non conforme au modèle de régression linéaire étant donnée l'étendue de la variation), mais on ne retrouve plus l'effet du contexte noté précédemment.

Ainsi, si le contexte prévocalique semblait favoriser *come* à la place de *com* plus fortement que le contexte préconsonantique dans les emplois *a priori* réservés à *com*, ce n'est plus le cas ici : les moyennes des taux de formes monosyllabiques sont proches (devant #C, moy. = 67% et moy. pond. = 60% __#V ; devant #V, moy. et moy. pond. = 62%), et l'effet mesuré par le test de Student est extrêmement faible ($t = 1,88$) et non significatif ($p = 0,06$). Il faut néanmoins noter, sur le graphique en boîtes à moustaches, un trait commun entre ces données et celles restreintes à la précession d'un pronom personnel sujet ou d'un verbe conjugué : c'est devant initiale vocalique que la distribution est la plus étalée, tandis qu'elle se concentre plus fortement autour de la médiane devant initiale consonantique.

La perte du contraste selon le contexte phonologique entre les données de cette section et celles de la section précédente peut être due à l'absence de <i> parmi les initiales vocaliques repérées par nos requêtes, cette exclusion ayant fait chuter le nombre d'occurrences repérées devant initiale vocalique alors même que le contexte syntaxique n'est plus un critère restrictif : devant pronom personnel sujet ou verbe conjugué, 2970 occurrences ont été repérées devant initiale vocalique, tandis qu'elles ne sont plus que 1787 dans le même contexte phonologique mais sans considération de la partie du discours du mot suivant. Cette sur-représentation de <i> initial après « come » tient très probablement à la domination des sujets masculins à la troisième personne dans la littérature médiévale, l'exclusion des *il* et *ils* provoquant cette baisse du nombre d'occurrences relevées. Aussi, il n'est pas possible d'interpréter avec certitude cette absence d'effet, qui peut tenir à un biais de sélection dans nos données.

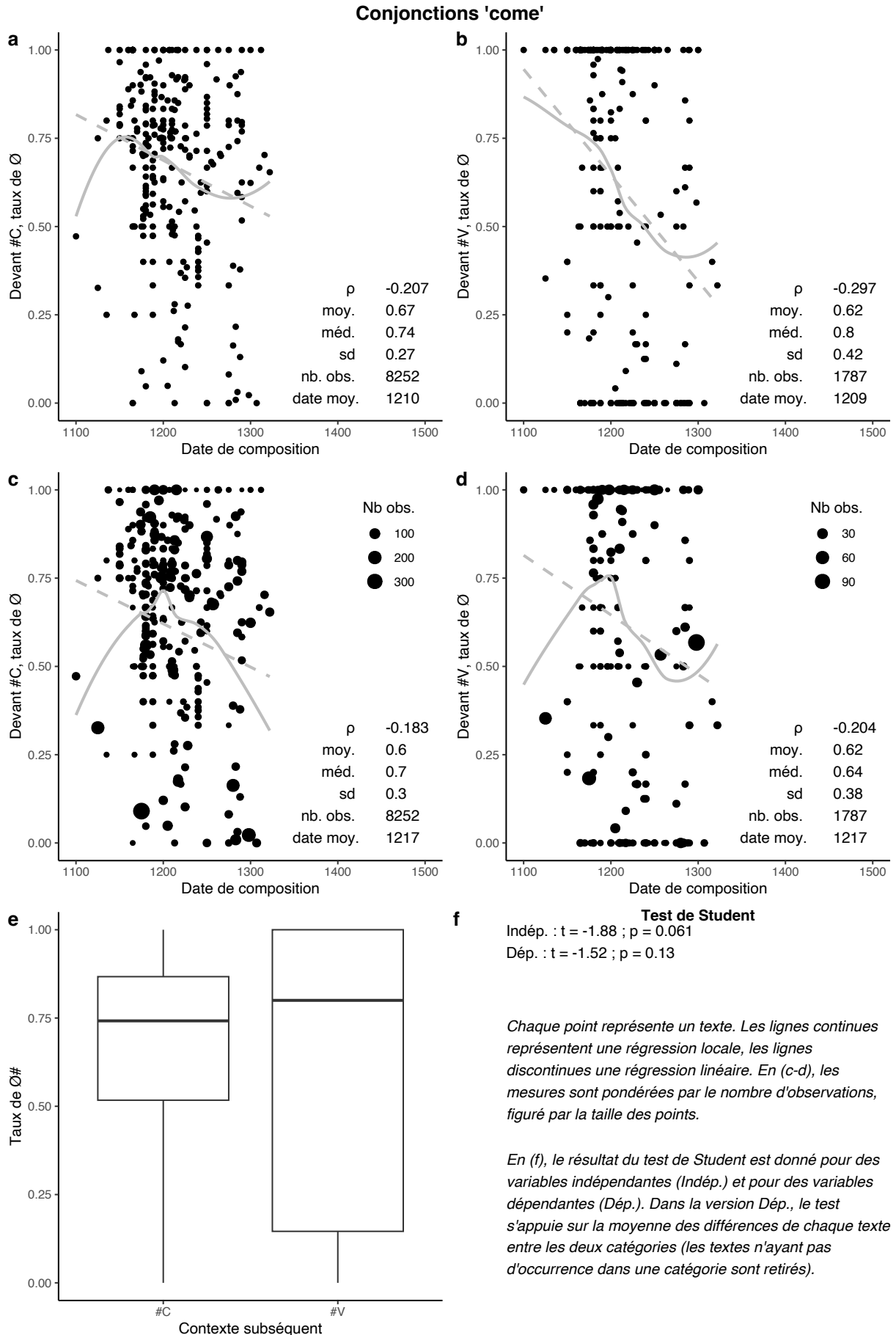


FIGURE 10.22 – Distribution contextuelle des variantes de conjonctions « come » en défaut de <e>.

À l'exception de cette disparition de l'effet du contexte, la situation d'ensemble est analogue à celle notée devant pronom personnel sujet ou verbe conjugué : le type *com* domine dans certains textes, dans d'autres c'est le type *come* qui domine, et cette domination du type *come* semble se renforcer avec le temps. Ce constat signifie que l'analyse proposée dans la section précédente, qui visait spécifiquement les contextes *a priori* dévolus au type monosyllabique, est au moins en partie valable pour l'ensemble des formes.

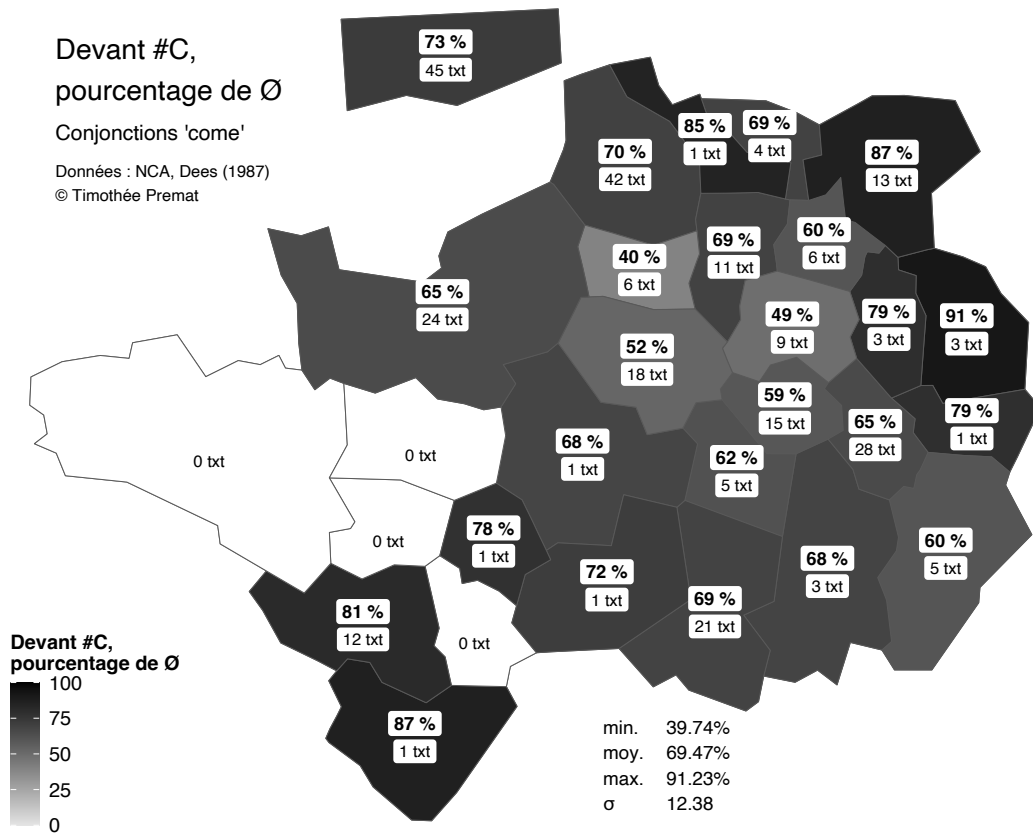
La comparaison entre boîtes à moustaches en fig. 10.18 (devant PROPERSONNEL ou VERBE) et celles en fig. 10.22 montre néanmoins une différence notable : alors que, devant #C comme devant #V, la médiane est à 100% devant pronom personnel sujet ou verbe conjugué, elle n'est plus qu'à 74% devant initiale consonnantique, et à 80% devant initiale vocalique dans tous les contextes syntaxiques. Corrolairement, les moyennes sont plus faibles dans tous les contextes syntaxiques (devant #C, moy. = 67%; devant #V, moy. = 62%) que devant pronom personnel sujet ou verbe conjugué (devant #C, moy. = 82%; devant #V, moy. = 68%). La différence majeure tient à la réduction du nombre de textes ne possédant que des formes monosyllabiques, particulièrement devant initiale consonnantique : moins d'un quart des textes n'ont que le type *com*.

Ainsi, le contrôle du contexte syntaxique induit une baisse de la quantité relative de formes du type *come*, ce qui indique que la distribution syntaxique joue bien un rôle dans l'ensemble des données, même si elle ne parvient pas à protéger le type *com* de manière systématique. Si l'effet de la distribution syntaxique est plus visible devant initiale consonnantique que devant initiale vocalique, il demeure néanmoins visible également devant initiale vocalique.

Distribution diatopique. En diatopie, la situation de l'ensemble des conjonctions « come » est analogue à celle observée en précession d'un pronom personnel sujet ou d'un verbe conjugué. Comme représenté en fig. 10.23, page suivante, c'est toujours le centre du domaine qui comporte le plus de *come* (donc les taux d'absence de <e> les plus faibles), et l'extension de ce centre ainsi que la puissance du remplacement de *com* par *come* est beaucoup plus réduite devant initiale consonnantique que devant initiale vocalique.

On note ainsi, devant initiale vocalique (fig. 10.23.b), trois régions centrales aux taux extrêmement bas (région parisienne : 15%, 16 textes; Yonne : 19%, 5 textes; Marne : 25%, 8 textes *vs* moy. = 67%); ces régions sont celles qui avaient les taux les plus bas devant pronom personnel sujet ou verbe conjugué à initiale vocalique (fig. 10.21.b *supra* : région parisienne : 11%, 17 textes; Yonne : 24%, 4 textes; Marne : 40%, 9 textes *vs* moy. = 74%). Devant initiale consonnantique (fig. 10.23.a), la région parisienne (52%, 18 textes) et la Marne (49%, 9 textes) demeurent dotées des deuxième et troisième taux les plus bas, mais ceux-ci sont sensiblement plus proches de la moyenne des taux régionaux (moy. = 69%). Le taux minimal dans ce contexte est atteint dans l'Oise (40%, 6 textes), et l'Yonne dispose cette fois-ci d'un taux plus proche de la moyenne (62%, 5 textes). Devant pronom personnel sujet ou verbe conjugué à initiale vocalique, région parisienne, Yonne, Oise et Marne disposent également de taux bas (resp. 57%, 15 textes; 63%, 5 textes; 66%, 5 textes; 70%, 9 textes *vs* moy. = 85%

a



b

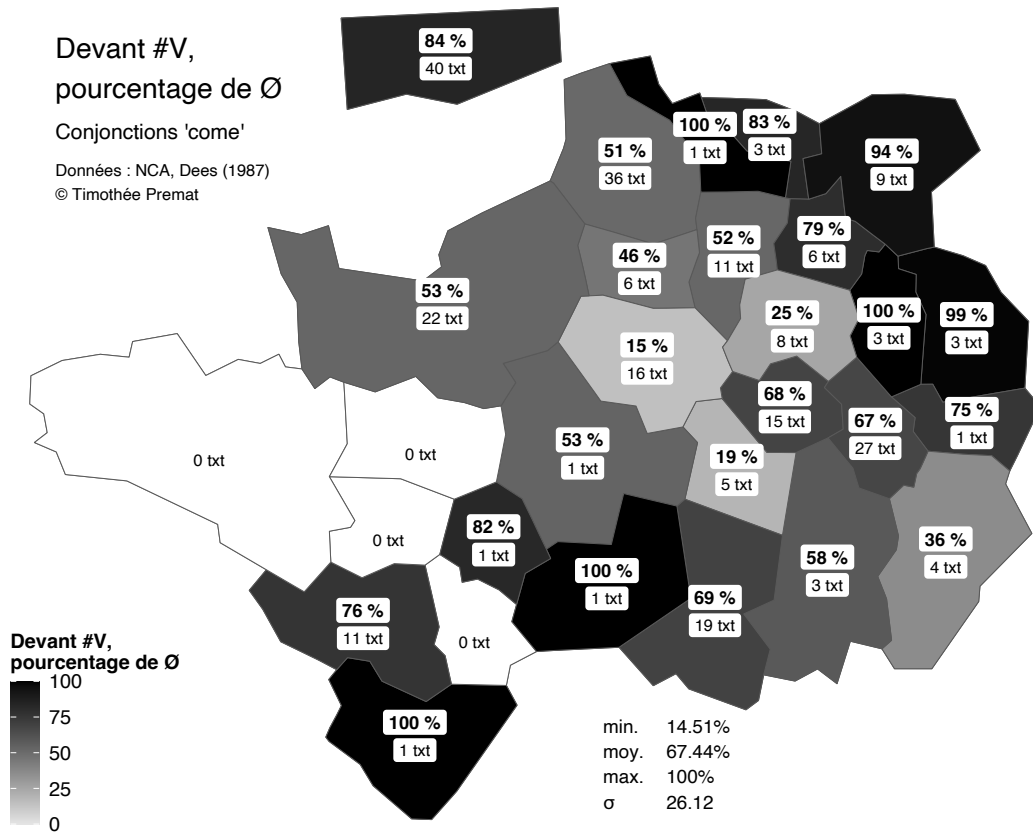


FIGURE 10.23 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, conjonctions « come » suivies d'une initiale consonnantique (a) et suivies d'une initiale vocalique (b).

en fig. 10.21.a *supra*). La superposition du pôle des taux les plus bas est donc partielle suivant les contextes syntaxiques et phonologiques ; cette superposition est résumée en table 10.5.

TABLE 10.5 – Superposition variable des taux moyens régionaux les plus bas de *com*, selon que le contexte syntaxique est contrôlé ou non, et selon l’initiale du mot suivant.

	PROPERSONS OU VER		Tous contextes	
	#C	#V	#C	#V
rég. parisienne	57%	11%	52%	15%
Oise	66%	58%	40%	46%
Marne	70%	40%	49%	25%
Yonne	63%	24%	62%	29%
Moy. rég.	85%	74%	69%	67%

On peut observer en table 10.5 la constance des taux bas de la région parisienne dans tous les contextes, syntaxiques comme phonologiques. En revanche, lorsque les contextes syntaxiques ne sont pas contrôlés, l’Yonne ne dispose d’un taux très bas (relativement à la moyenne des taux régionaux) que devant initiale vocalique et, devant pronom personnel sujet ou verbe conjugué, elle se démarque de la moyenne bien plus fortement devant initiale vocalique que devant initiale consonantique, tandis que l’Oise ne se démarque pas plus fortement dans ce contexte que dans les autres. Quant à la Marne, c’est devant pronom personnel sujet ou verbe conjugué à initiale consonantique qu’elle se démarque le moins de la moyenne des taux régionaux, disposant de taux assez faibles dans les autres contextes. La définition du noyau central dont les taux sont les plus faibles est donc variable selon le contexte syntaxique et phonologique, et seule la région parisienne a ce rôle nucléaire dans tous les contextes.

Intersection de la diatopie et de la diachronie. Les cartes en fig. E.20 - E.23, pages suivantes, représentent la diachronie de la distribution diatopique des alternances entre les formes du type *com* et celles du type *come*, selon les demi-siècles utilisés dans le reste de ce travail. Si, dans la première moitié du XI^e s., le centre du domaine n’est guère attesté étant donnée la rareté des textes continentaux (fig. E.20.a et E.22.a), il est attesté dans la seconde moitié du XI^e s. (fig. E.20.b et E.22.b), et dispose déjà des taux les plus faibles. Cette attestation est moins évidente devant initiale consonantique (fig. E.20.b) que devant initiale vocalique (fig. E.22.b), étant donnée l’extension plus restreinte de ce centre devant initiale consonantique.

Au cours du XIII^e s., le ‘centre’ identifié précédemment (en synchronie et dans la seconde moitié du XI^e s.) comme ayant les taux les plus faibles, semble s’élargir considérablement. Ainsi, la comparaison des cartes en fig. E.20.b et E.21.a montre, devant initiale consonantique, une baisse générale des taux à proximité de ce centre, qui atteint notamment l’Angleterre (qui passe de 76% à 56%), la Normandie (de 75% à 59%), l’Aisne (de 74% à 51%) et la Meuse (de 67 à 61%). Cette extension rapproche la distribution diatopique de l’axe Nord-Ouest - Sud-Est observée pour l’élément *que* et pour les conjonctions en *base* + *-que*. Si cette extension ne se poursuit pas après 1250 en termes d’augmentation de l’étendue

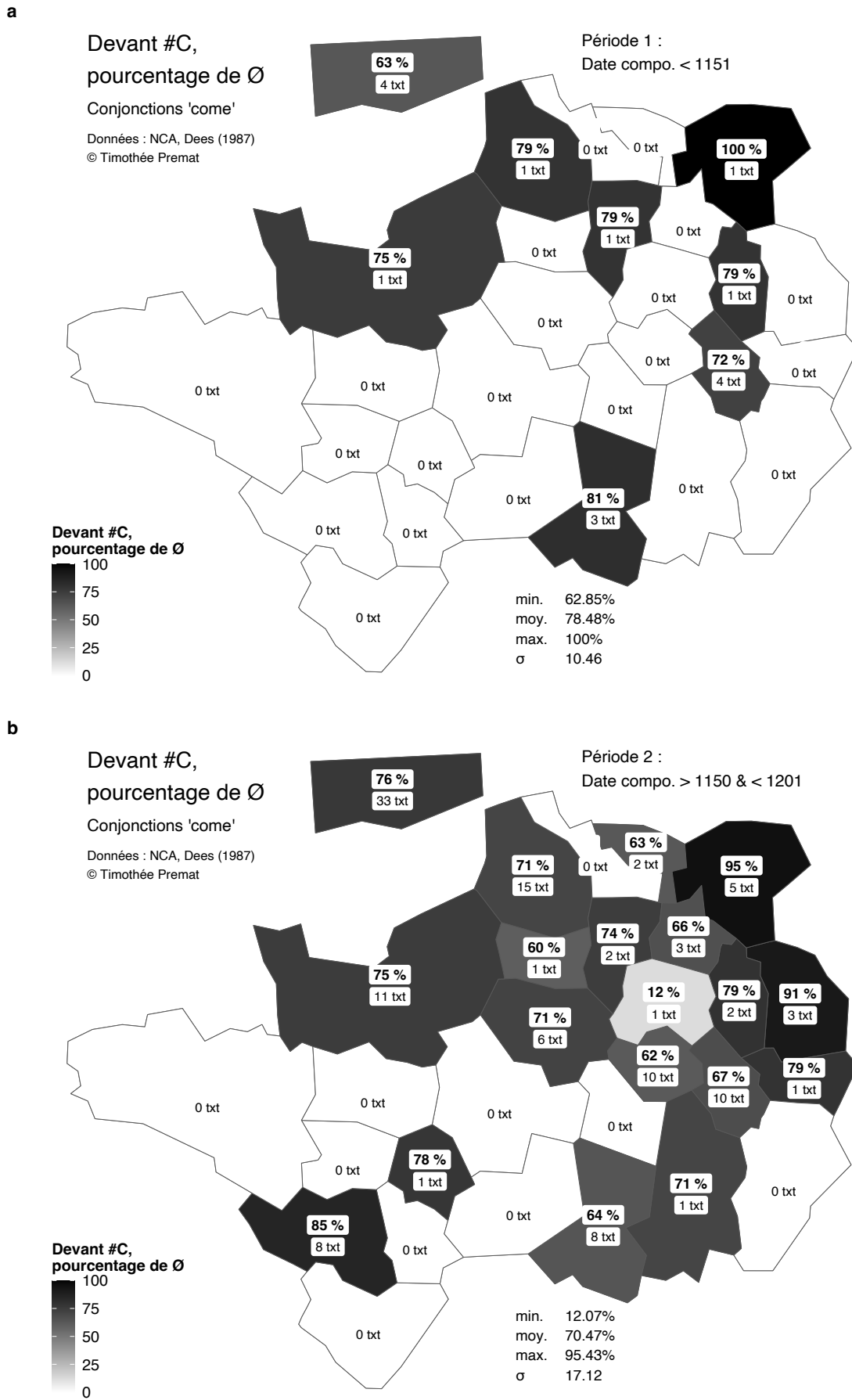
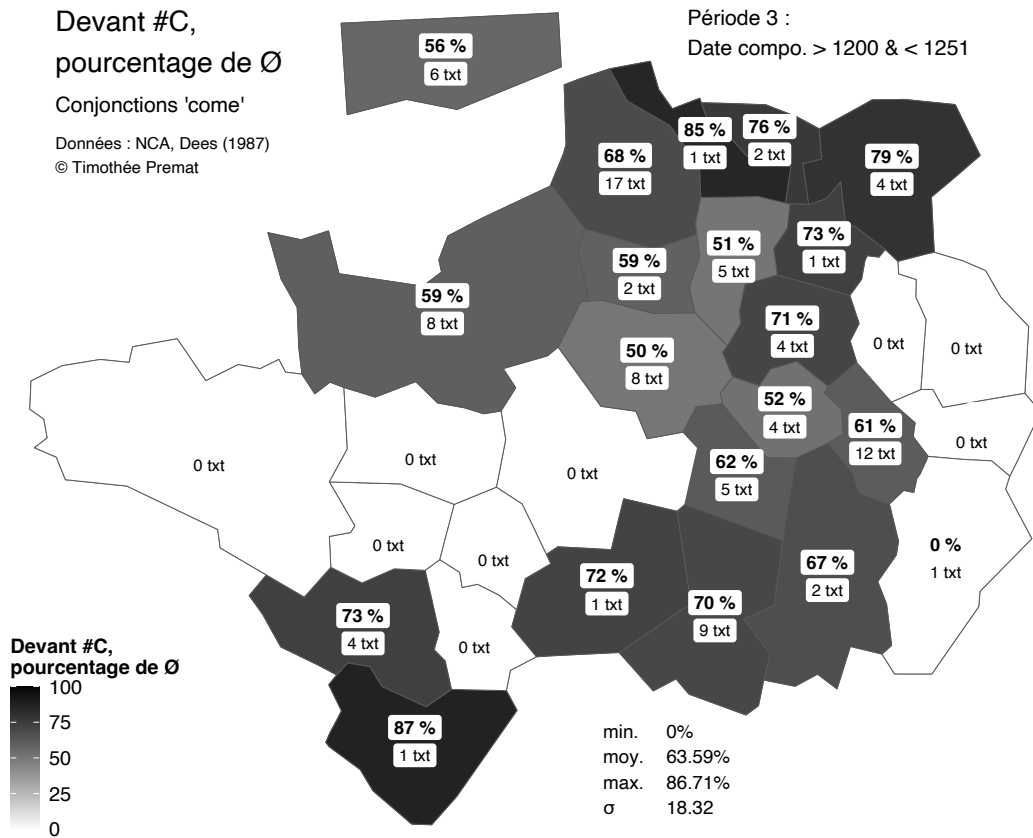


FIGURE 10.24 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1050) et 2 (b : 1151-1200), conjonctions « come » devant initiale consonantique.

a



b

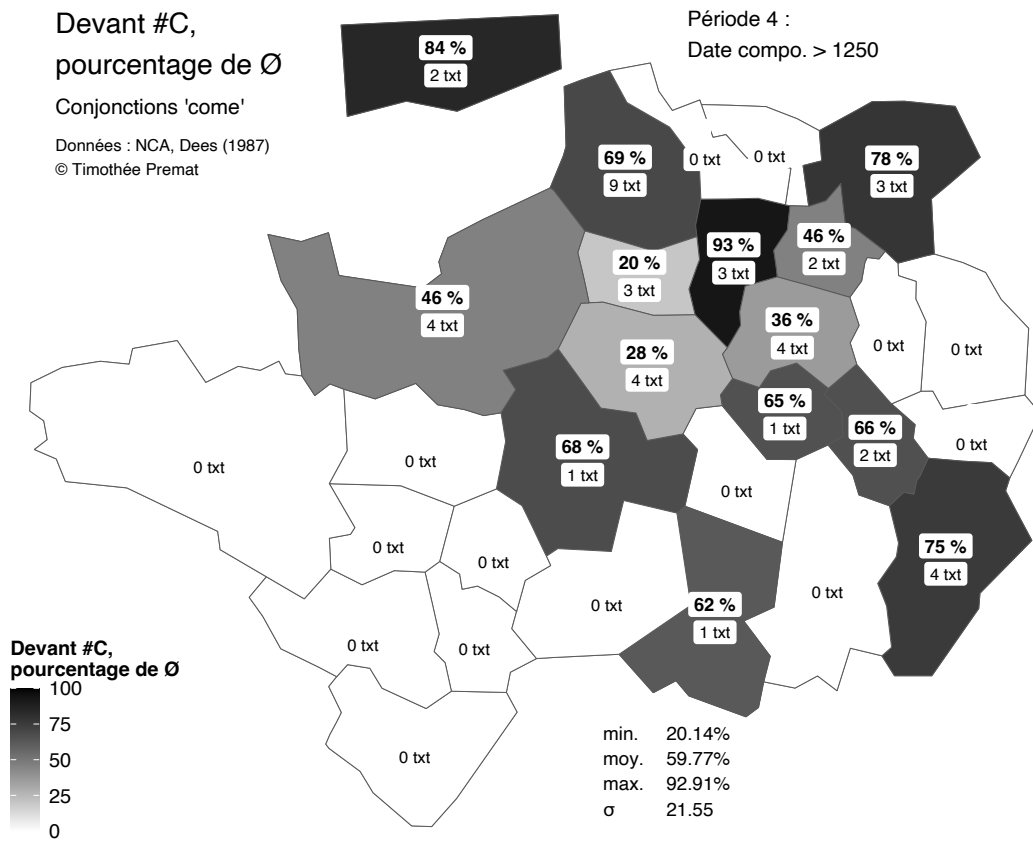


FIGURE 10.25 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), conjonctions « come » devant initiale consonantique.

spatiale, elle se poursuit en revanche si l'on observe les valeurs des taux : le centre du domaine s'affaïsse (région parisienne : de 50 à 28% ; Oise : de 59 à 20% ; Marne : de 71 à 36%), alors que les taux évoluent en général vers le bas, mais pas de manière aussi marquée (la moyenne des taux régionaux passe seulement de 64% dans la première moitié du ^{xiii}e s. à 60% après 1250).

Devant initiale vocalique, l'extension du domaine des faibles taux entre la seconde moitié du ^{xii}e s. et la première moitié du ^{xiii}e s. est encore plus marquée selon l'axe Nord-Ouest - Sud-Est mentionné, à l'exception notable de l'Angleterre (82%) et de l'Oise (85%). Ainsi, on observe dans la seconde moitié du ^{xii}e s. (fig. E.22.b) des taux très bas uniquement en région parisienne (33%) et dans l'Oise (20%), accompagnés par des régions plus ou moins continues au Nord et à l'Est dont les taux sont inférieurs à la moyenne des taux régionaux (Haute-Marne : 57% ; région Somme, Pas-de-Calais : 60% ; Aisne : 65% *vs* moy. = 70%). La situation change dans la première moitié du ^{xiii}e s. (fig. E.23.a), avec un effondrement des taux, de la Normandie (de 72 à 26%) et la région Somme, Pas-de-Calais (de 60 à 41%) jusqu'à la Bourgogne (de 100% à 38%) tandis que la moyenne des taux régionaux ne baisse que légèrement (de 71 à 66%). Enfin, après 1250, certaines périphéries conservent des taux très élevés (Angleterre et région Nièvre, Allier : 100% ; Wallonie et Ardennes : 75%), mais dont les plus élevés ne correspondent qu'à un seul texte (Angleterre et région Nièvre, Allier). Le reste des régions disposent de taux allant de 53% à 0% : on ne trouve plus aucune forme *com* en région parisienne, tandis qu'on trouve encore des formes *com* pour la moitié des occurrences en Orléanais et dans la région Somme, Pas-de-Calais. Les limites de la diagonale Nord-Ouest - Sud-Est ne peuvent être précisément établies, étant donnée l'absence de données après 1250 dans la plupart des régions marginales ; néanmoins, les Ardennes, la Wallonie, l'Angleterre et la région Nièvre, Allier — même si leurs taux sont à prendre avec précaution étant donné le faible nombre de textes — semblent témoigner de l'existence de cette "vallée" diagonale.

En somme. Les données sur la variation entre les types *com* et *come* sans filtrage en fonction de la partie du discours du mot subséquent sont en partie conformes à ce que nous avons observé pour « come » suivi par un pronom personnel sujet ou un verbe conjugué. Le mouvement général enregistré par ces données n'est donc pas seulement un mouvement de disparition de la distribution syntaxique, mais également un mouvement plus général dans la langue, en faveur du type bisyllabique. Par ailleurs, le contrôle du contexte syntaxique témoigne d'une certaine persistance de l'évitement du type *come* devant proposition dont le verbe est exprimé : on trouve proportionnellement moins de *come* devant pronom personnel sujet ou verbe conjugué que dans n'importe quel contexte, même si la comparaison est biaisée par l'inclusion des <i> à l'initiale des pronoms personnels. La distribution diatopique de ces formes est également assez sensible à celles devant pronom personnel sujet ou verbe conjugué, avec un centre où *come* domine plus fortement qu'en périphérie, et l'étude de la diachronie de cette diatopie confirme une influence qui émanerait d'un centre et se répandrait à partir de ce centre, de manière radiale et de préférence au Nord-Ouest et au Sud-Est.

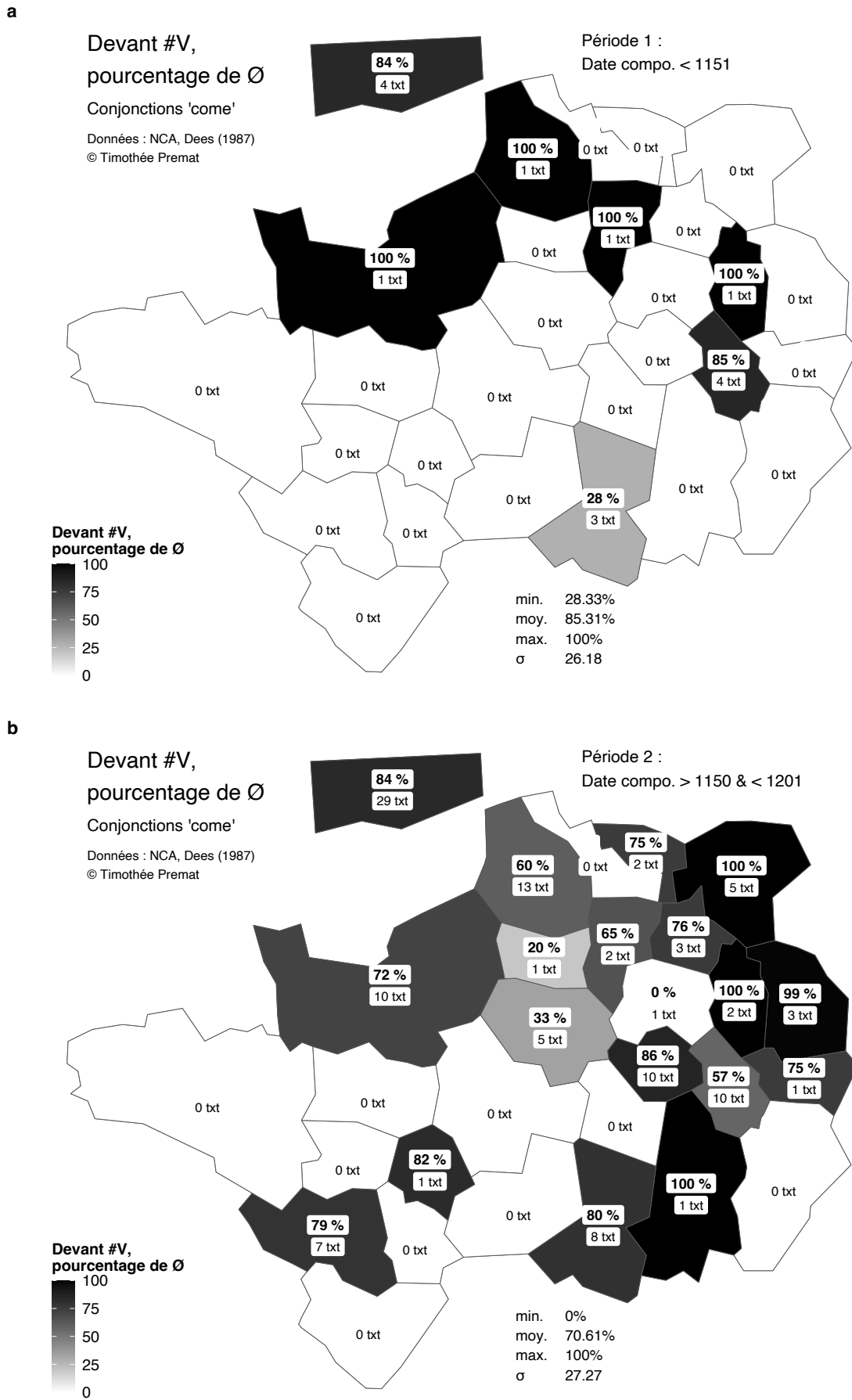


FIGURE 10.26 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1050) et 2 (b : 1151-1200), conjonctions « come » devant initiale vocalique.

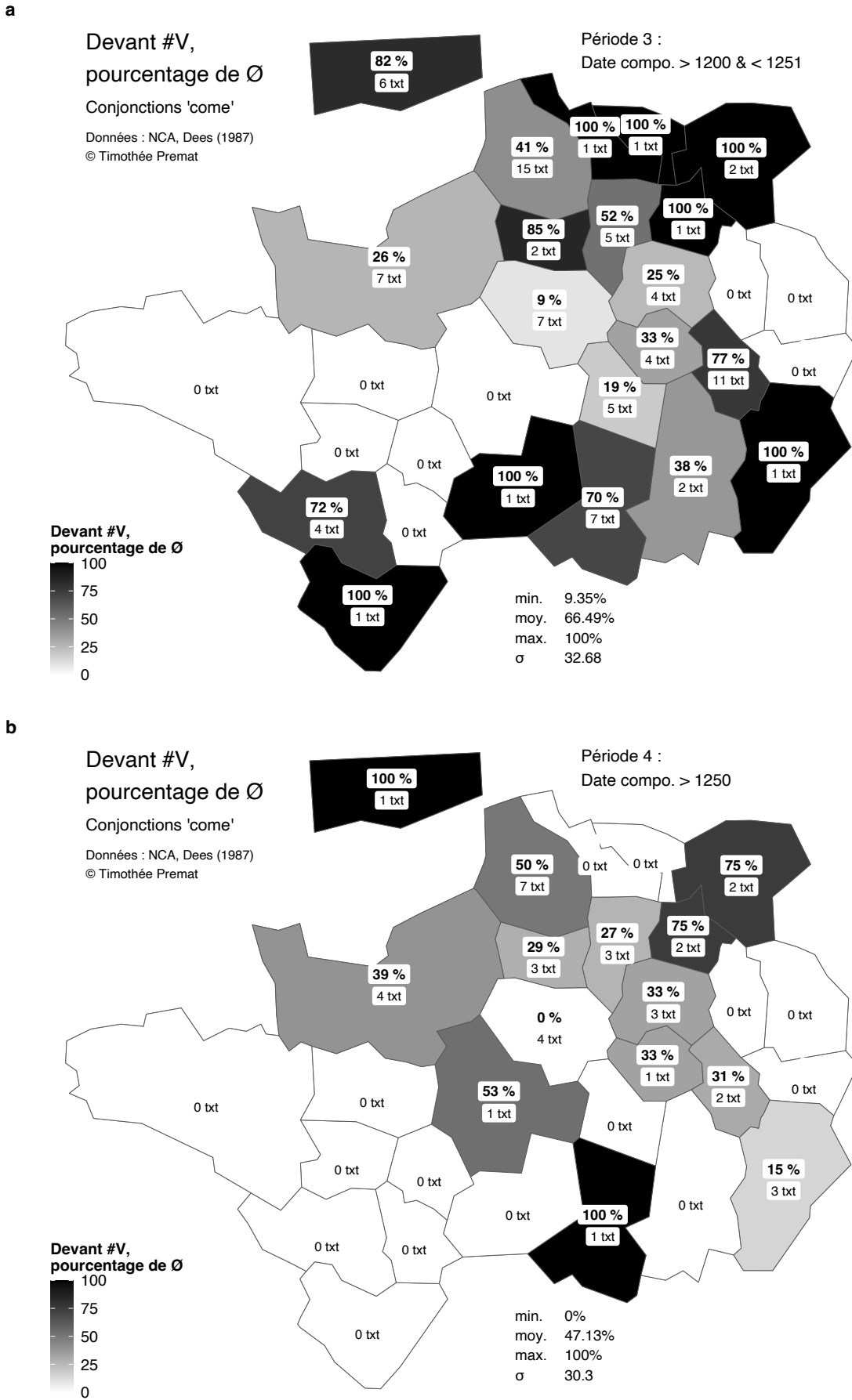


FIGURE 10.27 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), conjonctions « come » devant initiale vocalique.

10.2.4.3 En somme

La distribution syntaxique complémentaire entre les formes monosyllabiques du type *com* et celles bisyllabiques du type *come* semble s'effacer, dans certains textes, dès la moitié du ^{xiii} s. La dynamique est assez semblable selon que l'on ne considère que les contextes syntaxiques où c'est le type monosyllabique qui est attendu (devant pronom personnel sujet ou verbe conjugué) ou que l'on ne filtre pas les données en fonction de ces contextes. Dans les deux cas, on trouve de plus en plus de formes bisyllabiques, et une distribution des données plus étalée devant initiale vocalique que devant initiale consonantique. En diatopie, les formes bisyllabiques sont les plus fréquentes au centre du domaine, et ce centre est plus restreint devant initiale consonantique que devant initiale vocalique. À l'intersection entre diatopie et diachronie, on observe une contamination progressive à partir d'un noyau central, qui aboutit, après 1250, à la formation d'une diagonale Nord-Ouest - Sud-Est de taux bas, bien plus marquée devant initiale vocalique que devant initiale consonantique, mais qui ne comprend jamais l'Angleterre.

Aussi, la disparition de la distribution syntaxique complémentaire entre le type monosyllabique *com* et le type bisyllabique *come* est parallèle au mouvement de régression des formes monosyllabiques, peu importe le contexte syntaxique. Pour autant que les deux jeux de données soient parallèles, la raréfaction du type *com* au profit du type *come* est moins marquée devant pronom personnel sujet ou verbe conjugué, ce qui semble indiquer que ce contexte syntaxique induit une certaine résistance à la généralisation du type bisyllabique. Il n'y a donc pas lieu de penser que c'est la disparition de la distribution syntaxique qui entraîne la raréfaction progressive des formes monosyllabiques, mais plutôt de considérer que c'est le mouvement de généralisation progressive des formes bisyllabiques qui emporte la distribution syntaxique.

Ce mouvement semble tendre au remplacement de *com* par *come* au centre du domaine, voire sur un axe Nord-Ouest - Sud-Est, mais il ne concerne que faiblement les régions périphériques. L'unification des formes par la généralisation du type *come* semble donc être une innovation qui se répand à partir du centre du domaine, à la fois de façon radiale mais également selon un axe diagonal. De ce fait, la conservation, aux marges du domaine, de la domination du type monosyllabique *com* en introduction d'une proposition dont le verbe est exprimé n'est pas réductible au maintien d'une distribution complémentaire de type syntaxique, mais elle collabore avec une meilleure rétention des formes monosyllabiques aux marges du domaine, condition nécessaire à l'expression de cette complémentarité syntaxique. Ce n'est pas seulement la préférence pour le type *com* devant proposition non elliptique qui est mieux conservée aux marges du domaine, mais également les formes du type *com* en général.

10.3 Des observations aux généralisations

Dans cette section, nous proposons de résumer la profusion de données exposée *supra* à des principes généraux et qui soit sont analytiques soit sont idoines à servir de support à un discours analytique. Nous commençons par synthétiser ce que nous avons pu relever au niveau du

comportement distinct des lemmes ou plutôt, ici, des familles de lemmes (sec. 10.3.1), avant de revenir sur l'effet du contexte sur la génération de formes en défaut de <e> (sec. 10.3.2), puis sur l'évolution diachronique (sec. 10.3.3) et sur la distribution diatopique (sec. 10.3.4) des données. Enfin, nous terminons en replaçant nos six lemmes dans l'ensemble des conjonctions, pour envisager l'étendue de la variation au sein de la catégorie (sec. 10.3.5).

Nous ne dédions pas de section spécifique à l'anglo-normand, sa particularité étant abordée dans les sections sus-mentionnées. Nous ne traitons pas non plus de la question de l'influence de la fréquence sur les alternances <V>/Ø, étant donné que considérer les conjonctions en *base + -que* comme un ensemble cohérent reviendrait à n'avoir que trois catégories de lemmes (*base + que*, élément *que* et « come »), et qu'isoler les lemmes individuels reviendrait à travailler sur des fréquences trop faibles pour être significatives (moins de 3000 occurrences pour 30 lemmes).

10.3.1 Lemmes atteints par le défaut de <e>

Nous avons retenu trois familles de lemmes pour nos analyses : les conjonctions en *base + -que*, l'élément « que » et la conjonction « come ». Les conjonctions en *base + -que* sont au nombre de 30, et représentent près de 3000 occurrences, tandis que nous avons plus de 90 000 occurrences de l'élément *que* et un peu plus de 10 000 occurrences de « come ». Concernant les conjonctions annotées *pos* = "CON", aucune autre conjonction simple que « come » ne se termine par un <e>, et les alternances sont ici circonscrites à <e>/Ø (nous n'avons pas d'occurrences d'alternances <V>/Ø). Quant aux pronoms conjonctifs ou conjonctions annotées *pos* = "PROCON", si nous n'avons étudié que l'élément *que*, deux autres lemmes connaissent une variation <e,V>/Ø : *se/si* et *ne/ni*. Ce sont, avec l'élément *que*, les seuls lemmes annotés PROCON à se terminer par une voyelle finale atone. Aussi, tous les lemmes susceptibles d'avoir une variation <e,V>/Ø sont concernés par cette variation. Il ne semble donc y avoir, ici, aucune sélectivité lexicale : tous les items se terminant par un schwa (pour les conjonctions pures) ou par un schwa ou une voyelle finale atone et non initiale³¹ (pour les PROCON) sont susceptibles d'une alternance <e,V>/Ø. Ces éléments sont repris en (10.24), où le critère de voyelle finale non initiale est synthétisé par la présence d'une consonne avant la voyelle.

31: Cette précision est nécessaire pour exclure les cas où le PROCON se réduit à une seule voyelle, à la fois initiale et finale, tels *o*, *e*, etc.

(10.24) CONJ : sélectivité lexicale.

Tous les lemmes conjonctifs en (-)Cə ou (-)CV, susceptibles de connaître le défaut de schwa, le connaissent.

La situation de « come » est spécifique du point de vue syntaxique, l'emploi de *come* étant originellement réservé à l'introduction d'une proposition elliptique. Cette distribution complémentaire est assez bien conservée aux marges du domaine mais s'effondre rapidement à partir du centre du domaine (10.25).

(10.25) CONJ : sélectivité syntaxique.

La conjonction « come » répond à une répartition syntaxique archaïque, qui se perd au centre du domaine et se répand ensuite depuis ce centre.

De fait, les alternances <e>/∅ de « come » ne se comportent pas comme les alternances <e>/∅ des locutions en *base + -que* et les alternances <e,V>/∅ de l'élément *que* : d'une part le <e> de *come* n'est pas prédictible à partir de l'étymologie — si peu assurée soit-elle — et des règles de la phonétique historique, et d'autre part les taux du type monosyllabique *com* sont proches de ceux des conjonctions en *base + -que* devant initiale vocalique, mais « come » n'est pas sensible au contexte, alors que les conjonctions en *base + -que* et l'élément *que* ne connaissent que l'élision.

10.3.2 Le rôle du contexte

L'effet du contexte dépend fortement du lemme : l'élément *que* et les conjonctions en *base + -que* ne connaissent que l'élision (on ne trouve de défauts que devant initiale vocalique), tandis que « come » connaît des alternances <e>/∅ équivalentes devant initiale consonantique et devant initiale vocalique. Les taux d'élision de l'élément *que* et des conjonctions en *base + -que* sont néanmoins loin de la systématisme observée, par exemple, pour les articles *le* et *la* (99% de l'__#V), avec un taux moyen d'élision par texte de 69% pour les conjonctions en *base + -que* et de 41% pour l'élément *que*. Si la différence entre ces deux catégories peut provenir de notre incapacité à séparer les formes pronominales des formes purement conjonctives, il convient néanmoins de noter que l'élision n'est qu'optionnelle dans les deux cas. Enfin, pour « come », l'absence de sensibilité au contexte rapproche ce lemme du comportement des adverbes (cf. sec. 7.2.4). Ces éléments sont synthétisés en (10.26) et (10.27), dont les mesures proviennent des fig. 10.4, 10.11 et 10.22.

(10.26) CONJ : contexte, *base + -que* et élément *que*.

Le contexte prévocalique est le seul à être associé à des défauts de <e>, et cette association n'est pas systématique.

(*base + -que* : $t = 26$; $p < 0,005$; moy. __#V = 69%)

(élément *que* : $t = 32$; $p < 0,005$; moy. __#V = 41%)

L'effet n'est pas stable dans le temps et l'espace.

(10.27) CONJ : contexte, « comme ».

Les défauts de <e> ne semblent pas sensibles au contexte phonologique.

($t = -2$; $p = 0,06$; moy. __#V = 62%; moy. __#C = 67%)

L'effet n'est pas stable dans le temps et l'espace.

10.3.3 Diachronie générale

Si l'on observe des tendances générales sur les graphiques, celles-ci ne sont qu'approximatives, ne sont jamais conformes à un modèle de régression linéaire, et dépendent en réalité d'une forte interférence diatopique. Ainsi, la tendance à l'augmentation de l'élision des conjonctions en *base + -que* et de l'élément *que* que nous avons notée en sec. 10.2.2.1 et 10.2.3.1 n'est due qu'au poids de l'anglo-normand, qui pratique moins l'élision que le domaine continental tout en étant précoce (sec. 10.2.2.3 et 10.2.3.3). Si l'on isole les données continentales, la tendance est plutôt à la régression de l'élision graphique, mais cette tendance est très légère et est une nouvelle fois très loin de la conformité à un modèle de régression linéaire, du fait

de l'éclatement constant des données sur une large partie de l'échelle des ordonnées.

Pour « come », la tendance est également à la décroissance des taux d'absence de <e>. Si cette tendance n'est pas non plus conforme à un modèle de régression linéaire, elle est en revanche conforme au mouvement général qui mène, à terme, à l'élimination des formes monosyllabiques *com*. Si cette situation recoupe une nouvelle fois des distributions diatopiques contrastées, elle n'est cependant pas influencée outre-mesure par l'anglo-normand, dans la mesure où les textes anglo-normands pratiquent très peu le remplacement de *com* par *come* — contrairement à ce que pourrait laisser penser la lecture des remarques de SUCHIER (1879) et VISING (1895).³²

(10.28) **CONJ : diachronie.**

Il n'y a pas de réelle tendance diachronique pour l'élision des conjonctions en *base + -que* et de l'élément *que*. Pour « come », la tendance est à l'évacuation progressive du type *com* au profit du type *come*, dans tous les contextes syntaxiques.

32: Ces deux auteurs ne tiennent pas de discours général sur la situation de l'anglo-normand, mais mentionnent les premiers textes anglo-normands où la distinction syntaxique ne tient plus.

10.3.4 Diatopie

Pour nos trois catégories, on trouve une distribution diatopique remarquable : les omissions du <e> des conjonctions en *base + -que*, de l'élément *que* et de « come » sont plus faibles sur une diagonale allant du Nord-Ouest au Sud-Est, à l'exclusion de l'Angleterre pour « come ». Cette 'vallée' centrale où <e> (ou <i> dans le cas des formes animées de l'élément *que*) se conserve est plus marquée pour les conjonctions en *base + -que* que pour l'élément *que*, et se double d'une dynamique *centre - périphérie* pour « come ». Malgré ces nuances, l'existence de cette entité gréographique continue pour nos trois catégories est remarquable.

(10.29) **CONJ : diatopie.**

Les finales atones sont significativement plus maintenues sur une diagonale Nord-Ouest - Sud-Est qu'aux marges de celle-ci.

10.3.5 Place de la variation au sein des CONJ

Toutes les mesures citées *supra* concernent les défauts de <e> des occurrences des 32 lemmes sujets à défaut de <e>. Comme indiqué en sec. 10.3.1, les 30 lemmes en *base + -que* ainsi que « come » sont les seules conjonctions annotées comme telles dans le corpus qui sont susceptibles d'une variation <e>/∅, dans la mesure où l'alternance des conjonctions ne concerne pas d'autres voyelles finales que schwa et où ce sont les seules à se terminer en schwa. Aussi, peu importe la place de ces 31 lemmes au sein de la catégorie des conjonctions, toutes les conjonctions qui réunissent le critère élémentaire de se terminer par un schwa final sont sujettes à ces alternances. Il est en revanche à noter que d'autres mots annotés *pos* = "PROCON" que l'élément *que* sont également sujets à l'élision, même s'ils n'ont pas fait l'objet d'une étude poussée ici : c'est le cas de *si/se* et de *ni/ne*. Aucun autre lemme annoté PROCON ne se termine par une voyelle précédée d'une consonne (p.ex. « ou », « nen »),

ce qui confirme que l'élision est au moins optionnelle pour tous les éléments potentiellement conjonctifs se terminant soit par <e> pour les conjonctions pures soit par <e> ou <i> pour les éléments potentiellement pronominaux. Enfin, si les conjonctions en *base + -que* ne sont pas d'une fréquence très élevée (près de 3000 occ. une fois nos filtres contextuels appliqués), ce n'est pas le cas de « come » et de l'élément *que*, qui sont dotés d'une fréquence très élevée et ne peuvent donc être considérés comme mineurs en chiffres absolus. Ces éléments sont synthétisés en (10.30).

(10.30) **CONJ : poids de la variation.**

Les alternances avec Ø concernent toutes les CON qui se terminent par un schwa (a) et tous les PROCON qui se terminent par <e> ou <i> (b). Les défauts sont fortement marqués (c).

L'élément *que* et « come » ont une fréquence absolue élevée (d).

- a) CON : les 31 lemmes retenus contiennent 100% des lemmes en (-)Cə;
- b) PROCON : les 3 lemmes sujets à alternance <e,i>/Ø concernent 100% des lemmes en -CV;
- c) *base + -que* 0% __#C 69% __#V
 élément *que* 0% __#C 41% __#V
 « come » 67% __#C 62% __#V
- d) élément *que* : F = 90 250 occ.
 « come » : F = 10 039 occ.

10.4 Conclusion

Dans ce chapitre, nous avons étudié la distribution des défauts de <e> de certaines conjonctions dans le corpus du NCA. Après un bref point sur la situation des conjonctions en ancien français (sec. 10.1.1), nous avons relevé dans l'introduction que cette étude concernait les mots annotés comme des conjonctions (pos= "CON") ainsi que l'élément *que*, pronom conjonctif ou conjonction (pos = "PROCON"; sec. 10.1.2), et que les formes *com* de la conjonction « come » ne correspondaient pas à ce que nous caractérisons comme des défauts de <e> (sec. 10.1.3). Nous avons ensuite décrit le traitement des données (sec. 10.1.4), avant de présenter les résultats de l'analyse que Comparalem nous a permis de mener (sec. 10.2), et avons proposé une brève typologie des lemmes et formes sujets à défauts de <e> (sec. 10.2.1). Nous avons ensuite divisé la présentation selon cette typologie, en commençant par le sort du <e> des locutions en *base + -que* (sec. 10.2.2), avant d'aborder celui des voyelles finales de l'élément *que* (sec. 10.2.3) et de la conjonction *come* (sec. 10.2.4). Enfin, nous avons procédé à une montée en généralisation, qui tient lieu de *discussion* des résultats observés (sec. 10.3).

Cette généralisation nous a permis de réduire les faits observés à un ensemble de caractéristiques majeures. Au niveau des lemmes (sec. 10.3.1), nous avons noté que les lemmes concernés étaient les seuls lemmes de conjonctions pures à se terminer en /ə/ et que les deux autres lemmes annotés PROCON se terminant par une voyelle finale atone étaient également sujets à des variations <V>/Ø (10.2.4); la variation concerne donc ici tous les lemmes qui répondent à la simple condition phonologique

de se terminer par un schwa ou une autre voyelle finale atone. L'analyse contextuelle a révélé des propriétés différentes en fonctions des lemmes (sec. 10.3.2). Si le défaut n'est jamais systématique devant initiale vocalique pour les conjonctions en *base* + *-que* et pour l'élément *que*, il ne se produit jamais devant initiale consonantique, ce qui signale que c'est bien le processus contextuel de l'élision qui est à l'œuvre (10.26). En revanche, le contexte n'a pas d'effet sur les alternances *com/come*, qui se comporte sur ce point comme les adverbes (10.27).

Si nous n'avons pas noté (sec. 10.3.3) de tendance diachronique robuste pour les locutions en *base* + *-que* et pour l'élément *que*, nous avons en revanche noté que l'on percevait dans notre corpus la tendance au remplacement progressif du type *com* par le type *come* (10.28). L'analyse diatopique (sec. 10.3.4) a mis en lumière une distribution commune à nos trois catégories conjonctives, et ce malgré la diversité de leurs traitements de la finale atone : c'est au centre du domaine et sur un axe Nord-Ouest - Sud-Est que les défauts sont les plus rares, tandis que les finales atones sont mieux préservées aux marges de ce centre et de cet axe (10.29). Enfin, nous avons noté que nos cas de défaut de <e> étaient loin d'être marginaux au sein de l'ensemble des conjonctions (sec. 10.3.5). Si les défauts de <e> ne concernent en définitive que deux 'mots' (d'une part « que », autonome ou devenu suffixe conjonctif, et d'autre part « come »), les occurrences traitées sont fort fréquentes et représentent l'ensemble des conjonctions qui pourraient, phonologiquement, être sujettes à une variation <e>/∅, tandis que, pour les éléments annotés pos = "PROCON", les deux autres éléments se terminant par une voyelle sont également sujets à des alternances <e,V>/∅ (10.30).

De fait, les données de ce chapitre se distinguent de celles des autres chapitres : si l'on considère que le *-que* des locutions conjonctives est assimilable à l'élément *que* autonome, ce chapitre ne porte que sur deux lemmes, dont l'alternance de l'un, « come », n'est pas caractérisée comme procédant d'un défaut de <e>. Ce chapitre repose néanmoins sur un nombre considérable d'occurrences, et nous avons expliqué que cette restriction était apparemment due au hasard (les autres lemmes de conjonctions pures ne se terminent pas en schwa). Les données de ce chapitre sont également singulières dans la mesure où ce que nous avons appelé l'élément *que*, de loin le plus fréquent, recoupe des formes pronominales et des formes conjonctives, l'annotation du corpus ne permettant pas de distinguer entre les formes élidées de *que* et celles de *qui*, ni entre les occurrences purement conjonctives et celles qui sont pronominales.

De ce fait, il est permis de penser (mais cela n'est pas démontré en l'état) que le taux moindre d'élision de l'élément *que* procède de formes pronominales, et que les taux de défaut de *que* purement conjonctif doivent être analogues à ceux du suffixe conjonctif des locutions en *base* + *-que*. Mais il est également possible de postuler que cette différence provient du statut de lexicalisation des locutions en *base* + *-que*, qui pourrait expliquer l'écart vis-à-vis de *que* seul. Cette dernière hypothèse est contre-intuitive, dans la mesure où, jusqu'à présent, nous avons toujours noté que les monosyllabes s'élidaient bien plus fortement que les polysyllabes ; si l'on veut maintenir ce paradigme analytique, il semble donc que la première hypothèse soit préférable, et que le *que* des locutions conjonctives soit à analyser comme un *que* autonome. Quant à « come »,

si ses formes en <e> ne sauraient être prédictibles à partir d'un étymon populaire *como*, la communauté de distribution diatopique qu'il partage avec les autres conjonctions est remarquable, et son comportement contextuel analogue à celui des adverbes permet de ne pas le considérer comme tout à fait isolé de l'ensemble des lemmes sujets à défaut de <e>, ce qui est possible si l'on considère que les deux représentations sous-jacentes, /kom/ et /komə/, sont à risque de fusion dès lors que leur distribution syntaxique complémentaire n'est plus préservée; fusion qui résulte en la disparition du type monosyllabique.

Chutes des schwas prétoniques

11.1 Introduction

Jusqu'ici, tous les chapitres de cette partie ont été dédiés à des cas de défaut de <e> ou d'autres voyelles atones en position finale. Or, tel que généralement bien attesté dans les grammaires, il existe deux processus qui, en ancien français, suppriment des schwas internes : la réduction des hiatus et la syncope. Il importe donc de comparer ce qu'il advient des voyelles finales atones avec ce qu'il advient des schwas internes, pour déterminer si tout ce que nous avons documenté précédemment ne tient qu'à un mouvement général de suppression des schwas, ou si ce qui se déroule en position finale est réellement spécifique.

La réduction des hiatus internes est le nom que nous donnons au processus par lequel un hiatus, contenant un schwa, se trouve réduit à la seule voyelle non schwa qui le compose, tandis que la syncope désigne, ici, la suppression d'un schwa et la resyllabification de son environnement consonantique, c'est-à-dire la suppression d'une syllabe dont le noyau est un schwa au niveau sous-jacent. La polymorphie de l'adverbe *véraiment* représentée en (11.1) permet d'illustrer ces deux phénomènes ; toutes ses formes sont largement attestées.

(11.1) Réduction du hiatus et syncope de l'adverbe *v(e)rai(e)ment*

- | | | |
|-----|------------------|--------------------------------|
| (a) | <i>véraiment</i> | |
| (b) | <i>véraiment</i> | réduction du hiatus |
| (c) | <i>vraiment</i> | syncope |
| (d) | <i>vraiment</i> | réduction du hiatus et syncope |

En diachronie longue, les deux processus sont actifs et résultent, sauf rétro-formation savante, en la disparition totale de leur cible ; d'où la forme *vraiment* en fr.mod. Sauf à considérer les proparoxytons savants du type *angele* pour *ange*, ces processus ne concernent que des schwas prétoniques : un schwa ne pouvant être tonique, et la réduction des proparoxytons de formation populaire étant déjà advenue, les schwas non finaux ne sont que prétoniques. La syncope et la réduction des hiatus dont nous traitons ici ne concernent donc que des schwas prétoniques, mais, dans le second cas, la voyelle avec laquelle le schwa forme un hiatus peut être la tonique du mot (p.ex. dans le cas classique de *mëur* → *mur*, fr.mod. *mûr*).

Ces réductions de hiatus et syncope sont évidemment conditionnées par leur environnement : la phonotactique est impliquée dans la définition même de la réduction des hiatus, et le résultat de la syncope doit pouvoir correspondre à une configuration bien formée selon la phonologie de l'ancien français. Néanmoins, ce conditionnement phonotactique n'est pas un effet de sandhi, fût-il un sandhi interne : la syncope ne peut être considérée comme résultant d'un meilleur agencement des mélodies vis-à-vis des positions, et efface au contraire des configurations non marquées au profit de configurations, certes acceptables, mais plus marquées.

Dans ce chapitre, il s'agit donc de documenter les cas de suppression de schwas prétoniques, de manière à déterminer si les cas de suppression de schwa final ou d'autres voyelles finales atones tiennent simplement à une suppression généralisée de ces voyelles. En l'état, la bibliographie seule ne permet pas d'obtenir un aperçu satisfaisant de la progression de ces phénomènes dans l'espace et dans le temps, comme le note (MORIN 2000, p. 200). Nous avons donc adopté une méthode analogue à celle que nous avons mise en place pour les schwas finaux, mais, étant donné qu'il ne s'agit pas là de l'objet principal de notre travail, nous nous limitons à des sondages textométriques. Pour chaque phénomène, nous avons identifié huit lemmes sujets à la suppression du schwa interne, et avons utilisé une version spécialement paramétrée de Comparalem sur ces huit lemmes.

Nous commençons, dans l'introduction, par établir une première approximation de ces deux phénomènes à partir de ce qu'en disent les grammaires de l'ancien français (sec. 11.1.1), avant d'appliquer notre analyse par sondage textométrique à la réduction des hiatus prétoniques (sec. 11.2) et à la syncope (sec. 11.3). Nous discutons ensuite ces résultats (sec. 11.4), pour aboutir à deux conclusions : d'une part, si la réduction des hiatus internes et la syncope partagent des caractéristiques communes (sec. 11.4.2.1), elles ne sauraient découler d'un simple processus phonologique unifié (sec. 11.4.2.2) ; d'autre part les cas de chute de voyelles finales atones documentées dans les chapitres précédents sont irréductibles à ce qui serait une tendance générale à la chute de schwa dont témoigneraient la réduction des hiatus internes et la syncope (sec. 11.4.2.3). Au terme de ce chapitre, nous considérons donc que les chutes de voyelles finales ne sont pas explicables uniquement par une tendance générale de la langue à éviter ces voyelles.

11.1.1 Origine et destin des schwas prétoniques

En position interne, le schwa médiéval peut provenir de /e, o, ɔ/ issus de E, O, AU non initiaux précédés d'un groupe *obstruante + liquide* (11.2 a-b) ; en l'absence de groupe *obstruante + liquide*, la chute systématique précède le t.a.fr. (11.2 c). Les schwas prétoniques peuvent également provenir d'un A non initial (11.2 d) ou d'un /e/ en syllabe initiale provenant de E (11.2 e), de A (11.2 f) ou de O (11.2 g) (SÉGÉRAL et SCHEER 2020, GGHF, p. 324, 337).

(11.2) Origine de schwa prétonique

- | | | | |
|-----|--------|----------------------|-------------------------------------------------|
| (a) | OL/e/ | † <i>quadrifúrcu</i> | → <i>carrefour</i> ¹ |
| | | NŪTRĪMĒNTUM | → a.fr. <i>norrement</i> |
| (b) | OL/o/ | PETROSĪLĪUM | → a.fr. <i>perresil</i> → fr.mod. <i>persil</i> |
| (c) | C.C/e/ | DORMĪTŌRĪUM | → <i>dortoir</i> , * <i>dor(m)etoir</i> |
| (d) | A | ORNAMĒNTUM | → <i>ornement</i> |
| (e) | #CE | VENĪRE | → <i>venir</i> |
| (f) | #CA | CABĀLLUM | → <i>cheval</i> |
| | | MĀTŪRUM | → a.fr. <i>mëur</i> → fr.mod. <i>mûr</i> |
| (g) | #CO | † <i>conúcula</i> | → a.fr. <i>quenoille</i> → <i>quenouille</i> |

En accord avec la pratique de toute la bibliographie, il faut ici dissocier selon que le schwa prétonique se trouve en hiatus avec une autre voyelle ou non, les deux configurations ne suivant pas le même type d'évolution :

MORIN (2000). « La variation dialectale et l'interdiction des suites *Voyelle + e muet* dans la poésie classique ».

SÉGÉRAL et al. (2020). « Phonétique Historique ».

1: En syllabe ouverte, ĩ, évolue en /e/ (SÉGÉRAL et SCHEER 2020, GGHF, p. 217) ; c'est ce /e/ qui devint schwa.

le schwa interne en hiatus est promis à une disparition générale, alors que le schwa interconsonantique nécessite un contexte phonologique précis pour pouvoir chuter.

11.1.1.1 Schwa interne en hiatus

En hiatus, un schwa prétonique peut être initial (11.3 a) ou interne (11.3 b). Au niveau des lemmes, le hiatus se produit le plus souvent avec la voyelle tonique (cf. 11.3 a-b), et la plupart des grammaires ne relèvent que cette situation. Parfois, néanmoins, les deux voyelles en hiatus sont elles-mêmes prétoniques. Ainsi, parmi les multiples exemples relevés par REGULA (1955, p. 79-82), une typologie se dégage : les mots contenant des hiatus prétoniques résultent majoritairement d'une composition nominale (11.3 c), ou d'une dérivation flexionnelle (11.3 d, e) ou suffixale (11.3 f, g). Suivant la nature de la dérivation, le hiatus peut être de type $V\text{ə}$ (11.3 c, e-g) et/ou $\text{ə}V$ (11.3 d, g) ; ces dérivations et compositions diffèrent également selon que le hiatus existe dans la base de la dérivation ou dans le premier terme de la composition mais y implique une tonique qui devient prétonique lors de la dérivation ou de la composition (11.3 c-e), ou qu'il a été formé par la rencontre de la base et du suffixe (11.3 f, et g : *ecuerie* (← *e(s)cui*er + *-erie*)² et *plaidoiërie* (← *plaidoi*er + *-erie*)).

2: *L'écurie* désigne d'abord la communauté des écuyers (TLFi).

(11.3) Configurations des schwas (et <a>) internes en hiatus

(a)	# (C) $\text{ə}\acute{V}$	<i>*aetáticu</i>	→ <i>éage</i>
(b)	VC $\text{ə}\acute{V}$	ARMĀTŪRA	→ <i>armëure</i>
(c)	V ə	Mots composés :	
		MEDIA + NÓCTEM	→ <i>miënnuit</i>
		<i>bé(e)</i> + <i>gueule</i>	→ <i>beegueule</i> ³
(d)	{ ə , a}V	INF 1 ^{er} g. :	<i>crëer</i> (« créer »)
			<i>despëechier</i>
			<i>empëechier</i>
			<i>sëoner</i> ⁴
			<i>rëuser</i>
(e)	V ə	FUT/COND 1 ^{er} g. :	<i>crierai</i> , <i>-oie</i> (« crier »)
			<i>prierai</i> , <i>-oie</i>
(f)	V ə	ADV <i>-e-ment</i> :	<i>joliment</i>
			<i>vraiment</i>
			<i>forcëment</i>
			<i>absolüement</i>
(g)	{ $\text{ə}\acute{V}$, V ə }	Autres suffixes :	<i>plentëureus</i>
			<i>ecuerie</i> ⁵
			<i>plaidoiërie</i>
(h)	{ ə , a}V	REDEMPŤŌNE	→ <i>rëancon</i> , <i>räencon</i> ⁶
		LĚŔPĀRDUS	→ <i>lëopard</i>
		CRĚĀTŪRA	→ <i>crëature</i>

3: « Qui a la gueule béante ». Tous les dictionnaires ne l'attestent pas ; le TLFi cite DU CANGE (1954 [1883-1887]) *Bëe-gueulle*, voir aussi GD « beegueulle ».

4: *Saoner* : « récuser des témoins », « par extension, refuser, rejeter, rebuter » (GD) ; construit sur *sëon* (REGULA 1955, p. 81), *saon* : « suspicion, reproche fait contre des témoins, récusation, cause de récusation » (GD).

5: Le TLFi donne *esquërie* et *escuerie*, où le hiatus a été scindé par un /j/, cf. *infra*.

6: Seconde forme : FEW « rëdëmpŤio ». <a> et <e> peuvent être intervertis. Également *raançon*. Cf. (11.8) *infra*.

Un seul des exemples de Regula où le hiatus n'implique pas la tonique peut être dérivé entièrement par l'évolution populaire depuis son étymon : REDEMPŤŌNE → *rëancon* (11.3 h). Dans ce mot, le ö de la troisième déclinaison des imparisyllabiques fait échapper le hiatus à l'accent, tandis que E initial est protégé de la syncope par sa position initiale (SÉGÉRAL et SCHEER 2020, GGHF, p. 325) et évolue en / ə / comme attendu (*id.*, p. 331). Le second E est protégé par sa coda (*id.*, p. 293) et résiste à la syncope ;

le hiatus est formé par la chute prévisible de /d/ intervocalique (*id.*, p. 415). Dans la forme issue du NOMSG, REDĒMPTĪO → *rĕance*, le processus est le même mais le hiatus est tonique et se range donc dans la catégorie principale.⁷ L'évolution CRĒĀTŪRA → *crĕature* montre un autre exemple de conditions dans lesquelles un étymon latin peut générer un hiatus interne prétonique, mais elle est au moins partiellement savante, le mot provenant du fond lexical du latin chrétien (REY 2012; FEW). L'étymon comporte lui aussi un /e/ initial, qui résiste à la consonification attendue en latin tardif (SÉGÉRAL et SCHEER 2020, GGHF, p. 205) puisqu'il est précédé d'une attaque branchante,⁸ et que, étant initial, échappe à la syncope et s'affaiblit simplement en /ə/ comme attendu. Peut-être à la suite de sa précession par /ə/, ou simplement à cause du caractère savant de la forme, /a/ ne s'affaiblit pas en schwa lui aussi, et le hiatus /ə.a/ de l'a.fr. est ainsi formé. Quant à *lĕopard*, son initiale suit la même logique : E initial ne chute pas (← /e/) et s'affaiblit en schwa. La seconde voyelle, /o/, demeure problématique, attendu que son affaiblissement en schwa est également prédit, mais comme pour *crĕature*, on retrouve la configuration */ə.ə/ qui ne semble pas acceptable. L'affaiblissement de la voyelle en schwa ou son évolution vers /e/ est par ailleurs attestée une fois que le premier schwa disparaît, cf. *liepard* en (11.5 a) *infra*. Pour ce mot, il faut également noter que son étymon latin LĒOPĀRDUS, issu du grec λεοπάραδος, est composé de LĒO (lat.) ← λέων « lion » et de PĀRDUS ← πάραδος « panthère, léopard », les deux termes étant des emprunts orientaux obscurs (REY 2012).⁹ Que la composition fût empruntée ou non, elle n'est pas devenue opaque, et ce lemme peut également être rangé dans la catégorie des composés nominaux (11.3 c).

La possibilité de formation de hiatus atones hors dérivation ou composition est donc existante, mais elle est très peu attestée dans l'ensemble des grammaires. De fait, hors du paradigme verbal du premier groupe (INF et FUT/COND) où la flexion génère de nombreuses formes à hiatus interne, le type əV̄ domine. Le type V̄ə ne pouvant être que final, əV est rarement endogène et, sinon, dépend d'une dérivation, et Və semble toujours provenir d'une dérivation, soit par déplacement de l'accent, soit par mise au contact d'un schwa avec la voyelle finale de la base ou du premier terme.

Peu importe leur catégorie, après une première période de stabilité, ces schwas sont condamnés à disparaître durant le Moyen Âge. Dans une écrasante majorité des cas, seule l'autre voyelle du hiatus subsiste (11.4).

(11.4) Réduction des hiatus internes

(a)	# (C)əV̄	<i>ĕage</i>	→ <i>age</i>
(b)	VCəV̄	<i>armĕure</i>	→ <i>armure</i>
(c)	Və	<i>miĕnuit</i>	→ <i>minuit</i>
(d)	{ə, a}V	<i>beegueule</i> <i>despĕechier</i>	→ <i>bĕgueule</i> ^f → <i>despescher</i>
(e)	Və	<i>joliement</i> <i>v(e)raiment</i>	→ <i>jolyment</i> ^a , <i>joliment</i> ^f 10 → <i>vrayment</i> ^g , <i>vrament</i> ^f , <i>verament</i> ^a
(f)	{əV, Və}	<i>platĕureus</i> <i>escuĕrie</i> <i>plaidoiĕrie</i>	→ <i>plantureus</i> → <i>escurie</i> → <i>pleydoyrie</i> ^g , <i>plaidorie</i> ^g
(g)	{ə, a}V	<i>rĕancon</i>	→ <i>ranchon</i> ^g

7: Les deux formes *rĕance* et *rĕancon* semblent parfois être considérées par les dictionnaires comme deux lemmes distincts : le paradigme imparisyllabique latin aurait été brisé en un doublon en a.fr. Ainsi, les dictionnaires GD(c) et DMF disposent d'une entrée pour *rĕance* et d'une autre pour *rĕancon*, mais l'AND et le TL les rangent sous le même lemme.

8: L'exclusion des /e/ atones précédés d'une attaque branchante n'est pas explicite chez SÉGÉRAL et SCHEER (2020, GGHF, p. 205), mais peut être déduite du fait que le groupe /Cj/ formé « est toujours hétérosyllabique » (*id.*, emphase de la source), ce qui exclut d'emblée les attaques branchantes, qui ne peuvent admettre un noyau vide après elle : */TR.j/.

9: CHANTRAINE (1968) donne l'emprunt dans l'autre sens, du latin vers grec, *contra* GAFFIOT (1934), BAILLY et EGGER (1935), TLFi et REY (2012); ERNOUT et MEILLET (2001 [1932]) ne se prononcent pas.

^a AND

^f FEW

^g GD et GDc

La source n'est pas indiquée dans le cas des formes qui sont largement attestées (*idem infra*).

10: Forme canonique : *jolivement*, cf. *infra*.

Dans certains cas, néanmoins, le schwa a laissé une trace, aux niveaux graphique et phonologique ou au seul niveau graphique. À côté de l'évolution tenue pour populaire qui transforme le hiatus /ə.V/ en diphtongue /jV/ (11.5 a), une évolution semi-savante conserve l'une des valeurs du graphème <e> en transformant le schwa en /e/ (11.5 b) (REGULA 1955, p. 180; ZINK 2013 [1986], p. 166), tandis que l'analogie, notamment dans les formes verbales, réécrit la base (11.5 c) (*id.*), éliminant ainsi le hiatus, par la suppression ou consonification du schwa ou par l'insertion d'une approximante. Parfois, la rétention ne semble être que graphique, notamment au sein des futurs et conditionnels du premier groupe, où elle est stricte (11.5 d). Mais, parmi les formes où le schwa ne disparaît pas, le type le plus représenté est probablement celui qui résout la tension du hiatus par l'insertion d'un graphème consonantique (11.5 e). Cette épenthèse est d'ailleurs suspectée dès avant la réduction des hiatus, par exemple dans la *Chanson de Roland* (11.5 f) (SHORT 2007, p. 94), et ne concerne pas uniquement les hiatus contenant un schwa (mais cf. *infra*) ou étant en position interne (11.5 g). En vers, ces épenthèses augmentent le nombre de rimes possibles au sein d'un même texte (11.5 h), ce qui ne clarifie pas leur statut phonologique.

(11.5) Résistance du schwa en hiatus interne

- | | | |
|-----|--------------------------|---------------------------------------------------------------------------------|
| (a) | /ə/ → /j/ ^z | <i>crëer</i> → <i>crier</i> (« créer ») |
| | | <i>lëopard</i> → <i>liepard</i> |
| (b) | /ə/ → /e/ ^z | <i>crëance</i> → <i>créance</i> |
| | | <i>sëance</i> → <i>séance</i> |
| (c) | Analogie ^f | <i>vëez</i> → <i>voyez</i> ^z |
| | | <i>crëez</i> → <i>croyez</i> ^z |
| | | <i>pëestre</i> → <i>piestre</i> ^{r 11} |
| (d) | Rétention graphique | <i>crierai</i> , -oie (« crier ») |
| | | <i>prierai</i> , -oie ¹² |
| (e) | Épenthèse | <i>joliëment</i> → <i>jolivement</i> ^{a fg} |
| | | <i>escuërie</i> → <i>esquïerie</i> ^t , <i>escuierie</i> ^t |
| (f) | - archaïque ^s | <i>caiüz</i> pour <i>cäuz</i> |
| | | <i>seieler</i> pour <i>sëeler</i> |
| (g) | - {VV, Və#} ^s | <i>ahaise</i> pour <i>aaise</i> |
| | | <i>oyer</i> pour <i>oïr</i> |
| | | <i>raiendrat</i> pour <i>raendrat</i> |
| | | <i>espeye</i> pour <i>espee</i> |
| | | <i>nuwe</i> pour <i>nue</i> |
| (h) | - et rime ^s | /e.ə/ <i>entree</i> : <i>donnee</i> |
| | | /e.jə/ <i>entreie</i> : <i>veie</i> |

- a AND
f FEW
g GD et GDc
r REGULA 1955, p. 79-82
t TLFi
s SHORT 2007, p. 94
z ZINK 2013 [1986], p. 164-166

11: De PĒDĒSTER, l'adj. médiéval signifie d'abord « qui se promène sans but, qui vagabonde » (FEW). L'analogie, ici, fonctionne avec *pied* (REGULA 1955, p. 180). Un classement sous (11.5 a) aurait également été possible.

12: Nous n'avons trouvé aucune forme de FUT ou de COND où ces hiatus sont réduits.

Comme on peut l'observer en croisant les entrées en (11.5), ces différentes modalités de résistance graphique du hiatus peuvent se superposer, notamment quant au rôle de l'analogie, et le classement ici proposé n'est qu'imparfait. Par ailleurs, l'épenthèse n'est pas la seule mécanique de résolution du hiatus à concerner également les hiatus internes sans schwa : la réduction des hiatus peut également effacer des voyelles notées <a> (11.6 a) et, marginalement, des voyelles <o> (11.6 b) (ZINK 2013 [1986], p. 165).

(11.6) Réduction de hiatus sans <e> pour schwa (ZINK 2013 [1986], p. 165)

- | | | | | |
|-----|-----|------------|----------------|-----------------------------|
| (a) | <a> | †batáculat | → (il) baaille | → (il) baille ¹³ |
| | | CATÉNA | → chaeine | → cheine |
| | | CÁTHĚDRA | → chaiere | → chiere → chaire |
| (b) | <o> | CŎTŎNĚUM | → cooing | → coing |

13: *Bailler, baillir* : « Porter, emporter, donner, recevoir, louer, etc. » (GD, DMF, AND, FEW « bajulare ».)

Si les hiatus comportant un schwa et les hiatus ne comportant pas de schwa ne sont pas toujours distingués dans les grammaires, il y a là, pourtant, une question cruciale. Cette question est d'autant plus importante à soulever que la réponse ne semble pas évidente, dans la mesure où certains considèrent des assimilations de <e> pour schwa vers <a> ou <o> pour /a/ et /o/ comme étant une étape de la réduction des hiatus, et où les graphèmes vocaliques de ces hiatus sont, en réalité, peu stables. Ainsi, REGULA (1955, p. 79) propose pour certaines de ces voyelles une assimilation préalable à la chute (11.7).

REGULA (1955). *Historische Grammatik des Französischen*.

(11.7) Pour REGULA (1955, p. 79), l'assimilation précède la chute

- | | | | | |
|-----|---------|-----------|----------|------------------------------|
| (a) | ëage | → aage | → age | |
| (b) | mëaille | → maaille | → maille | |
| (c) | benëoit | → benooit | → benoit | → fr.mod. béni ¹⁴ |
| (d) | rëont | → roont | → ront | → fr.mod. rond |
| (e) | vëoir | → vooir | → voir | |

14: Mais cf. prénom *Benoît*.

L'assimilation qu'analyse Regula n'est pourtant pas certaine : dans la mesure où les grammairiens du xvi^e s. attestent d'un allongement compensatoire en cette position (ZINK 2013 [1986], p. 165), la double graphie <aa> ou <oo> peut très bien ne noter que cet allongement dès l'origine.¹⁵ De plus, les alternances entre <ae>, <ea>, <aa> et <ee>, voire <eo>, <ao>, <oo> et <oa>, sont nombreuses (11.8). De ce fait, ces graphies semblent témoigner d'un détachement de la graphie vis-à-vis d'une correspondance bijective avec les voyelles qui en respecterait la qualité.

15: Par ailleurs, on voit mal comment l'OCP peut admettre des séquences /a.a/ ou /o.o/ sans en provoquer leur transformation en /a:/ et /o:/, et, du moins à notre connaissance, ces séquences ne sont pas admises par ailleurs en a.fr.

(11.8) Graphies alternantes

- | | |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (a) | ëage, aage, eaage ^a , aegē ^a , adj. aegie ^f |
| (b) | mëaille, maaille, maele ^a , maaeye ^a |
| (c) | rëancon, rāencon, raancon ^{af} , reenceon ^f |
| (d) | rëont, raond ^a , roont ^{af} , rount ^{af} , ruunt ^{af} , roand ^a , rooant ^a , rooend ^a |
| (e) | benëoit et benëeit, benaieit ^a , beneieit ^a , benoois ^a |
| (f) | vëoir et vëeir, veer ^{af} , vaer ^a , veare ^a , vo(u)air ^{af} , voer ^{af} , vooir ^f |

^a AND
^f FEW

Ainsi, au vu des multiples façons d'écrire le hiatus ou l'ancien hiatus en (11.8), il semble difficile de soutenir l'hypothèse que les graphies <aa> et <oo> témoignent d'une assimilation intermédiaire entre /ə.a/ et /a:/ ou entre /ə.o/ et /o:/ . Le fait que le graphème <a> se mêle aux alternances entre <e> et <o> (*raond*, *roand*, *rooant* en (11.8 d)) renforce au contraire l'idée que ce graphème n'a plus la moindre valeur mélodique dans ce contexte, tandis que les graphies tri-vocaliques (*eaage*, *maaeie*, *rooant*, *rooent*) autorisent à penser que l'important, ici, est la notation de l'allongement compensatoire. Et ce d'autant plus que c'est en anglo-normand que ces notations tri-vocaliques sont attestées, or l'anglo-normand est particulièrement sensible à la notation de la quantité, et utilise déjà des doubles graphèmes pour ce faire (SHORT 2007, p. 51).

Par ailleurs, à la fin du Moyen-Âge, la même graphie hiatique peut librement correspondre à une ou deux positions métriques, et de nouveaux hiatus purement graphiques peuvent apparaître pour compenser la disparition d'un hiatus phonologique lors de la copie d'un vers, comme en atteste la graphie *chaasté*, dont la syllabe initiale correspond à deux positions métriques, qui est issue de *chasteé* après simplification du hiatus final (/fate:/) (MORIN 2000, p. 198). Ne pouvant écarter l'hypothèse que l'assimilation graphique ne représente que l'allongement, mais ne pouvant la démontrer non plus, nous la laissons de côté. Nous nous concentrerons donc sur l'analyse de la réduction des hiatus contenant un schwa, y compris lorsque celui-ci peut également être noté occasionnellement par un autre graphème que <e>, mais en ne prenant pas en compte la réduction de hiatus qui ne comporteraient jamais de <e> pour schwa.

11.1.1.2 Schwa prétonique hors hiatus

En syllabe prétonique et hors hiatus, le schwa échappe en principe à la syncope pendant le Moyen Âge (11.9 a) ; seul un ensemble restreint de mots, qui présente des contextes phonologiques plus ou moins réguliers, témoigne d'une syncope prétonique. Lorsque le schwa est initial, le contexte de syncope est nécessairement formé d'une obstruante (Obs) précédant immédiatement le schwa et d'un /r/ le suivant immédiatement (11.9 b) (SÉGÉRAL et SCHEER 2020, GGHF, p. 344 ; GESS, JACOBS et LAKS 2020, GGHF, p. 456). Ce contexte extrêmement restrictif permet aux deux consonnes de former une attaque branchante suite à la chute du schwa, ce qui évite la formation d'un groupe coda-attaque, prohibé en position initiale.

(11.9) Schwa initial hors hiatus

(a)		CABÁLLUM	→	<i>cheval</i> , * <i>chval</i>
(b)	#Obs__r	<i>verai</i>	→	<i>vrai</i> ^c
		<i>beroete</i>	→	<i>brouete</i> ^c
		<i>ferat</i>	→	<i>frat</i> ^s

^c SÉGÉRAL et SCHEER 2020, GGHF, p. 344

^s SHORT 2007, p. 95

En position prétonique interne, l'identification des contextes de syncope est plus complexe : pour NYROP (1899, p. 213) la syncope a surtout lieu après une liquide (LIQ) ou, éventuellement, une nasale (NAS ; 11.10 a),¹⁶ GESS, JACOBS et LAKS (2020, GGHF, p. 456, 478) et POPE (1966 [1934], p. 118) étendent ce constat au fait qu'au moins une des deux consonnes doit être liquide ou nasale (11.10 a-b), et les premiers notent (p. 478) que la syncope est plus fréquente entre une obstruante et une liquide. FOUCHÉ (1969 [1958], p. 514-515) précise que la syncope a lieu après /l/ et /r/ (11.10 a), entre une occlusive (OcCL) labiale ou vélaire et /l/ (11.10 c) et entre une occlusive ou une constrictive labio-dentale et /r/¹⁷ (11.10 d), tandis que ZINK (2013 [1986], p. 189) restreint la syncope aux contextes /n/__/r/ (11.10 a), /r/__/r/ (11.10 e) et, à partir du XIII^e s. en picard, à tous les contextes incluant un /r/. Ces formes syncopées ne remplacent pas immédiatement les formes non syncopées, et ces auteurs¹⁸ s'accordent sur le fait que la variation entre les deux types de formes perdure — jusqu'aux XVI^e et XVII^e s. pour FOUCHÉ (1969 [1958], p. 514).

16: NYROP (1899, p. 213) : la syncope a lieu « surtout après une liquide » ; ses exemples contiennent également des nasales.

17: FOUCHÉ (1969 [1958], p. 515) ne donne pas d'exemple avec une constrictive labio-dentale.

18: À l'exception de SÉGÉRAL et SCHEER (2020, GGHF, p. 344) et GESS, JACOBS et LAKS (2020, GGHF, p. 456, 478), qui n'en font pas mention mais semblent le sous-entendre.

(11.10) Schwa prétonique interne hors hiatus

- (a) {LIQ, NAs}__ sairement → serment^{NFZ}
 donerai → donrai^{NFZ}, dorrai^N
 derrenier → dernier^{PF}
- (b) __{LIQ, NAs} naperon → naprons^{PF}
 esperon → espron^{PF}
 souspeçon → soupçon^P
- (c) OcCL__ /l/ reguelisse → reglisse^F
 houbelon → houblon^F
- (d) OcCL__ /r/ acatera → acatra^Z
 demandera → demandra^Z
 esperit → esprit^F
 naperon → napron^F
- (e) /r/ __ /r/ durera → durra^Z
 plorera → plorra^Z

^F FOUCHÉ 1956, p. 514^G SÉGÉRAL et SCHEER 2020, GGHF, p. 344^N NYROP 1899, p. 213^P POPE 1966 [1934], p. 118^Z ZINK 2013 [1986], p. 189-190

Il y a donc, tel qu'indiqué en (11.10), de multiples définitions du contexte qui génère la syncope. Ces différents contextes ont pour seul trait commun de comporter au moins une liquide ou une nasale dans l'environnement immédiat du schwa promis à la syncope, c'est-à-dire une consonne dont la sonorité est élevée. On peut ainsi postuler que les contextes où le schwa promis à la syncope est entouré d'une obstruante et d'une liquide (11.10 c et 11.10 d) ou de deux /r/ n'est qu'une sous-catégorie de la catégorie générale des contextes déclenchant la syncope. La restriction de la syncope initiale aux seuls contextes *obstruante + liquide* (11.9 b) s'explique naturellement par l'absence de schwa en initiale absolue (non précédé d'une consonne) et par le fait que les groupes *coda - attaque* ne sont pas licites à l'initiale en a.fr. Dans tous les cas, ces contextes rendent la syncope licite : la syncope génère soit un groupe *coda - attaque* bien formé (11.10 a et 11.10 b), soit une attaque branchante (11.10 c et 11.10 d), soit ce qui a l'apparence graphique d'une géminée mais qui est selon toute vraisemblance réduit à une consonne simple (11.10 e).

Enfin, de manière distincte, certains mots connaissent une chute de schwas prétoniques internes entre /r/ et /v/ qui a lieu à époque pré-littéraire ; ces mots n'ont pas, ou très peu, de formes où la voyelle réapparaît ensuite. Ainsi, †*mirabilia* donne dès le très ancien français *merveille*, et les attestations de la forme non syncopée sont très rares : dans le NCA, nous n'avons trouvé que trois occurrences non syncopées (contre 1184 occurrences syncopées), à savoir deux occurrences *marevoille* dans un texte de la fin du xiii^e s. (pseuturp) et une occurrence dans un texte du début du xiv^e s. (cont re). Les mots du type de *merveille* sont donc à laisser de côté ; ils ont suivi un développement pré-littéraire idiosyncrasique, et l'apparition de rares formes non syncopées tient plus probablement à une épenthèse ou à une référence étymologique qu'à l'alternance réelle entre formes syncopées et non syncopées issues d'une même représentation mentale non syncopée.

Pour les schwas internes sujets à l'alternance <e>/Ø durant l'ancien français, la forme graphique n'est pas le seul témoin de l'apocope. Ainsi que le note SHORT (2007, p. 95), l'absence dans un vers d'un <e> graphique peut correspondre à l'absence d'une syllabe métrique, permettant de rétablir la bonne formation du vers, et ce dès le très ancien français. Ainsi, en (11.11 a) cité par SHORT 2007, p. 95, remplacer *durai* par *dunera*

pseuturp : ANON. (1290*). *The burgundian translation of the Pseudo-Turpin Chronicle*. Dees : Haute-Saône ; compo. bourg., ca 1290 ; ms. *idem* ; MS (WALPOLE 1949, 1948) ; DEAF : Turpin^{7W}

contre : ANON. (1322*). *Le Roman de Renart le contrefait, fin de la 6^e et dernière branche*. Dees : Aube ; compo. champ. mérid., ca 1322 ; ms. nil, ca 1337 ; CR2 (LEMAÎTRE et RAYNAUD 1914) ; DEAF : RenContr^{1R}

provoquerait un premier hémistiche de 5 syllabes, irréductible par la coupe épique (il est oxyton), et donc un vers de 11 syllabes métriques au lieu de 10. Le raisonnement peut être poussé plus loin : dans certains cas, la présence d'un <e> graphique dans un contexte d'apocope (tel que ces contextes sont listés en 11.10) semble ne pas correspondre à la présence d'une syllabe métrique. Ainsi, en (11.11 b), faire correspondre *trovereiz* à trois syllabes métriques crée un second hémistiche de 7 syllabes métriques et donc un vers de 11 syllabes métriques au lieu de 10, tandis que supposer que *trovereiz* représente ici /tro'vrejz/ permet d'obtenir une isométrie satisfaisante au niveau de l'hémistiche et du vers. Si l'on présuppose, et ce n'est certes qu'une présupposition, que le vers *doit* être correct, c'est là la correction du vers la plus économique, étant cohérente avec les conditions de chute du schwa interconsonantique. L'exemple en (11.11 b) est issu des résultats du PAM (cf. sec. 3.3), mais SHORT (2007, p. 95) cite les références de nombreux autres exemples. Les deux exemples cités fournissent déjà des éléments de datation : avec *Alexis*, le phénomène remonte au t.a.fr.

(11.11) Traitement métrique des schwas prétoniques hors hiatus

- (a) <Ø> = Ø *tut te durai, boens hom, quanque m'as quis*
 « Je te donnerai tout, homme bon, ce que tu
 m'as demandé »
 (*Alexis*, v. 224, trad. personnelle)
- (b) <e> = Ø *quer iloec est et iloc le trovereiz*
 « Car il est là, et vous le trouverez là »
 (*id.*, v. 315, trad. personnelle)

RAINSFORD et MARCHELLO-NIZIA 2018, *Vie de saint Alexis*

Si, évidemment, la métrique ne saurait être seule un témoin fiable d'un phénomène phonologique, la conjonction de cas où <e> pour schwa fait défaut et où le vers est isosyllabique (tels 11.11 a) et de cas où <e> pour schwa est présent et où le vers est hypermétrique (tels 11.11 b) plaide en faveur de la pertinence de la métrique, dans la mesure où la graphie atteste par ailleurs de la syncope. Et ce d'autant que, dans les deux cas, le contexte consonantique correspond bien à l'un des contextes propices à la chute de schwa interne.

11.2 Développement des chutes de schwa en hiatus əV

Si de nombreuses grammaires abordent la description de la réduction des hiatus internes, celles qui datent le phénomène aboutissent à des conclusions variées, ce qui appelle à établir une datation fiable, au sein de notre corpus, de ce phénomène. De plus, sur le plan diatopique, les indications sont rares et peu précises : ZINK (2013 [1986], p. 164) indique que la réduction n'est attestée qu'au Nord du domaine au XIII^e s. et ne se répand ailleurs qu'au XIV^e s., tandis que POPE (1966 [1934], p. 116) analyse simplement une plus grande instabilité de schwa dans les hiatus au Nord que dans le reste du domaine.

ZINK (2013 [1986]). *Phonétique historique du français*.

POPE (1966 [1934]). *From Latin to Modern French with Especial Consideration of Anglo-Norman*.

Par ailleurs, la réduction des hiatus où le schwa est précédé d'une voyelle atone (type Və) ne semble concerner que des cas de dérivation : pour la composition, nominale dans *miënnuit*, adverbiale dans *veraiëment*, et

pour la formation des futurs et conditionnels du premier groupe ((*jo*) *crierai*, *prierai*, etc.) (ZINK 2013 [1986], p. 165). Quant aux hiatus composés d'un schwa précédant une voyelle atone (type əV), ils semblent encore plus rares (*rëancon*), et parfois liés à des formations savantes (*crëature*). Aussi, pour disposer de suffisamment d'occurrences tout en évitant de ne représenter une catégorie que par un paradigme verbal, nous ne traitons ici que des configurations les plus courantes, où le hiatus est composé par la succession d'un schwa et d'une voyelle tonique (type əV̄ : MĀTŪRUM → a.fr. *mëur* → *mur* → fr.mod. *mûr*).

Cette section traite donc des chutes de schwas en hiatus avec la voyelle tonique suivante, en ajoutant au discours bibliographique des données portant sur un échantillon de lemmes sujets à la réduction du hiatus. Elle aborde d'abord la diachronie de la disparition de schwa dans cette position (sec. 11.2.1), et, après avoir exposé la méthode ici retenue (sec. 11.2.1.1), développe les premières attestations de réduction du hiatus (sec. 11.2.1.2) et le développement subséquent sous forme de co-existence de graphies réduites et non réduites (sec. 11.2.1.3). Elle aborde ensuite la distribution diatopique des réductions de hiatus (sec. 11.2.2) avant de traiter de l'intersection entre diatopie et diachronie (sec. 11.2.3).

11.2.1 Diachronie de la réduction des hiatus əV̄

Le moment précis où commence la réduction des hiatus əV̄ par chute du schwa n'est pas consensuel dans les grammaires, celles-ci proposant des datations qui vont du XII^e au XIV^e s. Ainsi, LA CHAUSSÉE (1989 [1974], p. 90, 209) les date du XII^e s., tandis que REGULA (1955, p. 80) et FOUCHÉ (1969 [1958], p. 520) les datent de la fin XII^e s.; POPE (1966 [1934], p. 110-111, 116), PRICE (1985 [1971], p. 76) et ZINK (2013 [1986], p. 164) les datent du XIII^e s.; É. BOURCIEZ et J. BOURCIEZ (1974 (1967), p. 44) les datent du XIV^e s.; MATTE (1982, p. 155) les date du XVI^e s. Certains auteurs donnent une période au sein de laquelle se déroule le processus, et insistent sur le moment où la suppression devient véritablement usuelle : LA CHAUSSÉE (1989 [1974], p. 90, 209) et FOUCHÉ (1969 [1958], p. 520) donnent une fourchette du XII^e s. au XV^e s., La Chaussée précisant que la réduction du hiatus est encore rare au XII^e s. et ne se généralise qu'au XV^e s. et Fouché précisant son *terminus ad quem* comme étant la seconde moitié du XV^e s.; POPE (1966 [1934], p. 110) donne une réduction du hiatus active jusqu'au XVI^e s. et indique que la réduction du hiatus devient usuelle au XIV^e s., É. BOURCIEZ et J. BOURCIEZ (1974 (1967), p. 44) indiquent que cette suppression se déroule à partir du XIV^e s.; PRICE (1985 [1971], p. 76), plutôt que de donner une fourchette, donne une alternative, la réduction s'effectuant aux XIII^e ou XIV^e s.; ZINK (2013 [1986], p. 164) indique que la réduction n'est attestée qu'au Nord du domaine au XIII^e s. et ne répand ailleurs qu'au XIV^e s., là où Pope ne notait qu'une plus grande instabilité, non chronologique, de ces schwas hiatus.

Enfin, certains auteurs distinguent variation phonologique et variation graphique. Ainsi, POPE (1966 [1934], p. 110) note que la graphie 'traditionnelle', c'est-à-dire hiatus, est retenue jusqu'au XVI^e s., en concurrence avec la graphie à hiatus réduit. De manière quelque peu analogue, ZINK (2013 [1986], p. 165) note que le schwa « amui, en precession, continue de s'écrire jusqu'à la fin du XVI^e siècle et au-delà pour certains mots. » Si la formulation de Pope ne présuppose pas que le hiatus est universellement

LA CHAUSSÉE (1989 [1974]). *Initiation à la phonétique historique de l'ancien français.*

REGULA (1955). *Historische Grammatik des Französischen.*

FOUCHÉ (1969 [1958]). *Phonétique historique du français — Les voyelles.*

PRICE (1985 [1971]). *The French language: past and present.*

É. BOURCIEZ et al. (1974 (1967)). *Phonétique française, Étude historique.*

MATTE (1982). *Histoire des modes phonétiques du français.*

réduit, Zink semble en revanche supposer que cette réduction est belle et bien advenue. Ces hypothèses jettent une lumière différente sur la période d'activité de la réduction du hiatus : si l'on s'en tient à l'analyse des formes graphiques, la présence de formes graphiques hiatiques n'indique en rien la survivance d'une variation phonologique entre hiatus non réduits et hiatus réduits et, à partir de la première occurrence réduite, on peut considérer que la réduction des hiatus est advenue. La paraphonologie métrique pouvant être, par définition, conservatrice, aucune preuve métrique ne peut invalider formellement cette hypothèse : la présence de nombreux <e> hiatiques maintenus et métrifiés ne saurait justifier de la réalité de ce maintien hors du discours versifié.

11.2.1.1 Méthode : sélection de huit lemmes hiatiques

Face à ces différences d'appréciation du phénomène, nous avons recherché dans le NCA les occurrences non hiatiques de lemmes du type ə́. Nous avons arrêté notre choix sur huit lemmes, suffisamment attestés dans le NCA ; ils sont donnés en (11.12).

(11.12) Formes hiatiques retenues

(a) <i>abbatissa</i> (lat.eccl.)	→ <i>abëesse</i>	→ <i>abesse</i>
(b) <i>ARMĀTŪRA</i>	→ <i>armëure</i>	→ <i>armure</i>
(c) <i>capillatūra</i> (b.lat.)	→ <i>chevelëure</i>	→ <i>chevelure</i>
(d) † <i>mercatāntem</i> (lat.pop.)	→ <i>marchëant</i>	→ <i>marchant</i>
(e) † <i>taetātūcu</i>	→ <i>ëage</i>	→ <i>age</i>
(f) <i>mire</i> ¹⁹ + <i>-ōir</i>	→ <i>mirëoir</i>	→ <i>miroir</i>
(g) <i>paradīsus</i> (lat.chrét.)	→ <i>pareīs, parais</i> ²⁰	→ <i>parvis</i> ²¹
(h) <i>peccatōrem</i> (b.lat.)	→ <i>pechëeur</i>	→ <i>pecheur</i>

Les formes correspondant à ces lemmes ont été obtenues par des requêtes lexicales dans TXM, soit uniquement par lemme lorsque le lemmatiseur était d'une précision satisfaisante, soit en collectant l'ensemble des variations graphiques rattachées au lemme et en cherchant des occurrences équivalentes où le schwa hiatique serait manquant. Cette méthode échappant à la lemmatisation impose une restriction sur le nombre de lemmes, dans la mesure où la recherche sur formes graphiques doit concerner des mots sans homographes.²² Ainsi, les formes du type *pareīs* ne peuvent être sujettes qu'à une recherche sur la base du lemme, leurs variantes *parais* notées *parais* se confondant avec l'INDPRÉ3sg du verbe « paraître ». Ces formes ont ensuite été données en entrée d'une version spécialement paramétrée du programme Comparalem, sur lequel repose notre analyse.

11.2.1.2 Premières attestations de la réduction des hiatus ə́

La première forme réduite de hiatus dans la configuration ə́ apparaît, dans notre corpus, à la fin du XII^e s., avec 9 formes *pecheur(s)* dans le texte anglo-normand en prose *psautier*.²³ Quatre autres lemmes apparaissent également avec un hiatus réduit au XII^e s., à commencer par le lemme « march(e)ant », dont *nimd* comporte une forme *marchans* et deux formes *marchant*, et dont *nimc* comporte une forme *marchant*. Ces deux textes sont des versions différentes du *Charroi de Nîmes*, texte versifié dont la

19: Base du verbe *mirer*.

20: Les formes *pareīs, parais* du lemme « paradis/parvis » posent en d'autres termes la question de la réduction des voyelles non notées <e> discutée *supra*, dont rien n'indique *a priori* que le graphique représente bien une voyelle spécifiée.

21: À l'exception des formes du type *paravis*. Les trois types de formes renvoient aussi bien au parvis d'une église qu'au paradis (TLFi « parvis »), et, par extension du premier sens, à des formes architecturales de clôture (AND « parei »). La formation populaire hiatique ou en /v/ provient de l'adaptation en /v/ du <δ> /ð/ du g. *παρόδεισος* (TFLi), tandis que la forme savante conserve un <d> intervocalique tout au long du Moyen Âge.

22: Dees ayant nettoyé préalablement ses textes, le NCA ne contient aucun des tréma que nous restituons ici pour distinguer hiatus et diphtongues.

psautier : ANON. (1125*). *Livre des psaumes*. Dees nil; compo. a.n., ca 1125; ms. a.n., ca 1150; CR2 (MICHEL 1876), DEAF : PsCambrM.

23: Toutes les occurrences en question sont au pluriel, échappant ainsi à l'alternance de base entre CSsg *pechiere* (← *peccātor*) et CRsg *pech(e)ur* (← *peccatōre*).

nimd : ANON. (1150*). *Le charroi de Nîmes*. Ms. D; Dees : Meuse; compo. nil, ca 1150; ms. lorr. mérid., ca 1275; MS1 (transcr. inédite L. Schoesler).

nimc : ANON. (1150*). *Le charroi de Nîmes*. Ms. C; Dees : Somme, Pas-de-Calais; compo. nil, ca 1150; ms. art., 1295; MS1 (transcr. inédite L. Schoesler).

composition remonte au milieu du ^{xii}e s. Apparaissent ensuite les formes à hiatus réduit *armure*, *age* et *chevelure*, avec respectivement une, une et deux occurrences, dans un texte anglo-normand de la seconde moitié du ^{xii}e s., *reis*. Ce texte semble particulièrement sujet à la pratique de la réduction des hiatus, puisqu'il dispose aussi de deux formes *marchant* et d'une occurrence *pecheurs* (CRpl); il ne dispose que de deux occurrences non réduites de nos huit lemmes, *marchèant*. Enfin, au ^{xiii}e s., apparaissent les formes à hiatus réduits des lemmes « par(e/v)is », « ab(e)esse » et « mir(e)oir ». La forme à hiatus réduit et consonne /v/ *parvis* apparaît d'abord au début du siècle dans *barril*, un texte de la région Nièvre, Allier selon la localisation de Dees (coef. dial. 85) ou orléanais selon les données philologiques, et dont c'est la seule forme à hiatus réduit. La forme *ab(b)esse* apparaît à 13 reprises dans *modw*, un texte anglo-normand de la première moitié du siècle, tandis que la forme *miroir* n'apparaît qu'au milieu du ^{xiii}e s. dans *fabb*, un recueil de fabliaux que Dees localise en Haute-Marne mais pour lequel la philologie donne une composition picarde et une copie bourguignonne. L'ensemble de ces premières attestations est résumé en table 11.1.

TABLE 11.1 – Premières occurrences de formes à hiatus əV réduits dans le NCA.

Lemme	Première occurrence			
	Date compo.	Date ms.	Région	ID
<i>Pech(e)eur</i>	1125	1150	Angleterre	psautier
<i>March(e)ant</i>	1150	1275	Meuse	nimd
		1295	Somme	nimc
<i>Arm(e)ure</i>	1175	1190	Anglettre	reis
<i>Chevel(e)ure</i>	1175	1190	Anglettre	reis
<i>(E)age</i>	1175	1190	Anglettre	reis
<i>Par(e)is</i>	1218	1317	Nièvre	barril
<i>Ab(e)esse</i>	1230	1250	Angleterre	modw
<i>Mir(e)oir</i>	1250	1310	Haute-Marne	fabb

Ainsi, notre recherche de huit lemmes dans le NCA confirme les datations les plus précoces, c'est-à-dire celles de LA CHAUSSÉE (1989 [1974], p. 90) et FOUCHÉ (1969 [1958], p. 520). En effet, les premières occurrences de 5 lemmes sur 8 datent du ^{xii}e s., et non des ^{xiii}e ou ^{xiv}e s. Néanmoins, la datation de Fouché semble être légèrement trop tardive vis-à-vis de notre corpus, dans la mesure où il précise que la réduction de ces hiatus ne commence qu'à la fin du ^{xii}e s. Même en nous fondant sur la date du ms. et non sur la date de composition, nos neuf premières occurrences *pecheur(s)* sont datées du milieu du ^{xii}e s. et non de la fin du siècle.

11.2.1.3 Cooccurrences des graphies <eV> et <V>

Comme suggéré par POPE (1966 [1934], p. 110) et ZINK (2013 [1986], p. 165), la graphie du hiatus réduit (consistant en un seul graphème vocalique) n'efface pas les graphies qui représentent le hiatus. Notre analyse avec le programme Comparalem montre que cette cooccurrence est très marquée en défaveur de la graphie réduite : non seulement la graphie réduite ne parvient pas à évincer la graphie bivocalique, mais en vérité elle demeure bien moins attestée que la graphie bivocalique durant toute la période de l'ancien français. Sur l'ensemble de la période considérée (de 1100 à 1322), seuls 16% des occurrences de nos huit lemmes disposent d'une graphie

reis : ANON. (1175*). *Les quatre livres des Rois*. Ms. M; Dees : nil; compo. a.n., ca 1175; ms. a.n., ca 1190; MS (prob. CURTIUS 1911).

barril : JOUHAM DE LA CHAPELE DE BLOIS (1218*). *Le conte dou barril*. Dees : Nièvre, Allier; compo. orl., ca 1218; ms. orl., ca 1317; MS (JOUHAM DE LA CHAPELE DE BLOIS 1932); DEAF : ChevBarBloisB.

modw : ANON. (1230*). *Saint Modwenna*. Ms. O; Dees : Angleterre; compo. a.n., ca 1230; ms. a.n., ca 1250; CR (A. T. BAKER et BELL 1947); DEAF : ModvB².

fabb : ANON. (1250*). *Fabliaux n°2, 4*. Ms. B; Dees : Haute-Marne; compo. pic., ca 1250; ms. bourg. sept., ca 1310; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).

réduite ; la moyenne des taux de chaque texte est moins élevée, avec seulement 9% de taux moyens de réduction des hiatus. En effet, sur les 214 textes qui répondent à au moins un de nos lemmes, 173 ne disposent que d'attestations à graphie hiatique, avec de 1 à 29 occurrences hiatiques (moyenne : 4 occurrences par texte). Et, parmi les textes qui présentent des graphies à hiatus réduit, seuls 9 ne disposent d'aucune graphie hiatique (de 1 à 10 graphies réduites par texte, en moyenne 4 occurrences par textes). Ainsi que figuré en 11.1, ces 9 textes sont distribués sur l'ensemble de la période, à partir de la première occurrence réduite et jusqu'à la fin de la période considérée. Enfin, 32 textes disposent de graphies réduites et de graphies hiatiques, avec en moyenne 32% de graphies réduites.

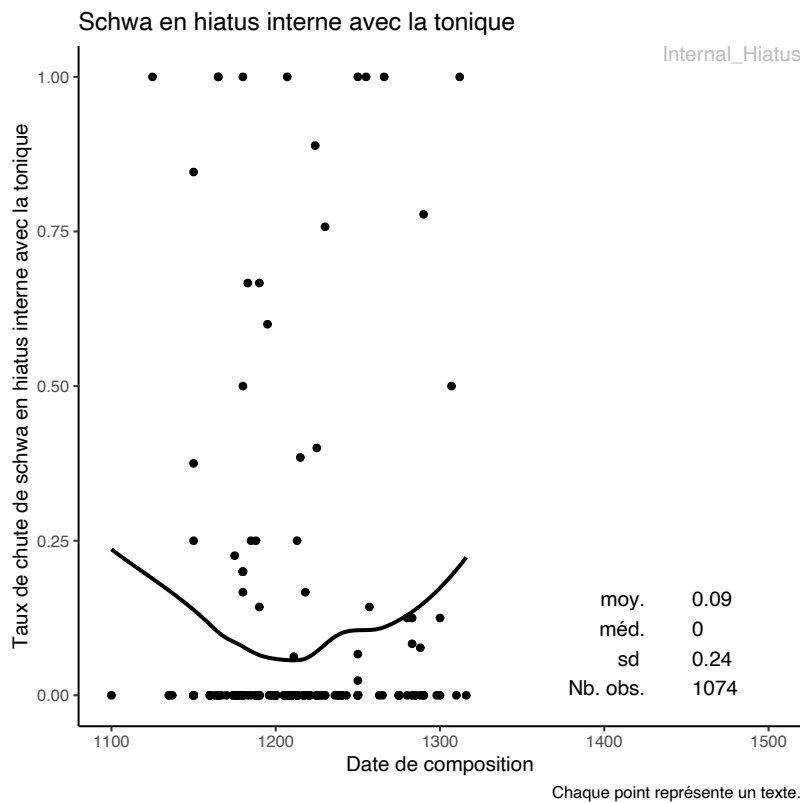


FIGURE 11.1 – Distribution diachronique des occurrences de hiatus réduits dans le NCA.

Pour chercher à équilibrer les valeurs entre les textes, de manière à ne pas sur-représenter des textes longs qui pratiqueraient l'une des deux graphies plutôt que l'autre, vis-à-vis de textes courts qui feraient l'inverse, il importe de regarder ce qu'il en est de la moyenne des textes et non de la moyenne des occurrences. De ce point de vue, les formes hiatiques l'emportent encore plus largement, avec une moyenne des taux de formes réduites qui descend à moins de 10%. La phase descendante initiale de la courbe de régression locale en fig. 11.1 n'est pas significative : elle est due au déplacement de la fenêtre mobile et aux deux textes dotés de taux de réduction des hiatus supérieurs à 70%. La seconde partie de la courbe, en revanche, est plus significative : elle dessine une progression régulière des graphies hiatiques, d'environ 10% au début du XIII^e s. à près de 25% au début du XIV^e s. En revanche, cette progression n'élimine pas la présence de textes n'appliquant aucune réduction des hiatus.

Ainsi, le panorama diachronique de la réduction des hiatus əV̄ peut être caractérisé comme suit : la réduction commence dès le XII^e s. et sa diffusion lexicale (son application à un lemme donné et pas à un autre) s'étend sur plus d'un siècle dans notre corpus. Si la graphie avec hiatus réduit était systématique une fois un lemme atteint par la réduction, on verrait en 11.1 une augmentation par paliers de 1125 (*pecheur* dans *psautier*) à 1250 (*miroir* dans *fabb*), dont chaque incrément correspondrait à un nouveau lemme atteint ; c'est loin d'être le cas, ce qui confirme une forte co-existence des deux graphies pour chaque lemme, à une période donnée voire, plus rarement, au sein d'un même texte. Ainsi, la graphie réduite <V> ne parvient jamais à remplacer la graphie bi-graphémique <əV̄>, qui demeure majoritaire sur l'ensemble de la période considérée, mais, tout en demeurant minoritaire, la graphie réduite semble prendre progressivement de l'ampleur au cours du XIII^e s.

11.2.1.4 En somme

Les premières attestations de réductions des hiatus sont précoces, datant de la première partie du XII^e s. Le phénomène, ou du moins sa transcription graphique, ne se répand dans le lexique que progressivement, avec des lemmes qui ne sont atteints qu'au milieu du XIII^e s. Si la graphie réduite peut être systématique au sein d'un texte, elle ne l'est jamais au sein des textes d'une période donnée, et ne semble progresser que dans la seconde moitié du XIII^e s., tout en demeurant minoritaire.

11.2.2 Diatopie de la réduction des hiatus əV̄

Sur le plan diatopique, les grammaires que nous avons consultées ne précisent que rarement le développement de la réduction du hiatus. Tout au plus, ZINK (2013 [1986], p. 164) indique que la réduction n'est attestée au XIII^e s. qu'au Nord du domaine, tandis que POPE (1966 [1934], p. 116) note que les graphies réduites sont plus courantes au Nord du domaine qu'ailleurs. Non seulement la datation de Zink n'est pas cohérente avec nos données, mais nous verrons également que la prééminence, chronologique ou statistique, donnée au Nord du domaine doit être reformulée en des termes plus précis.

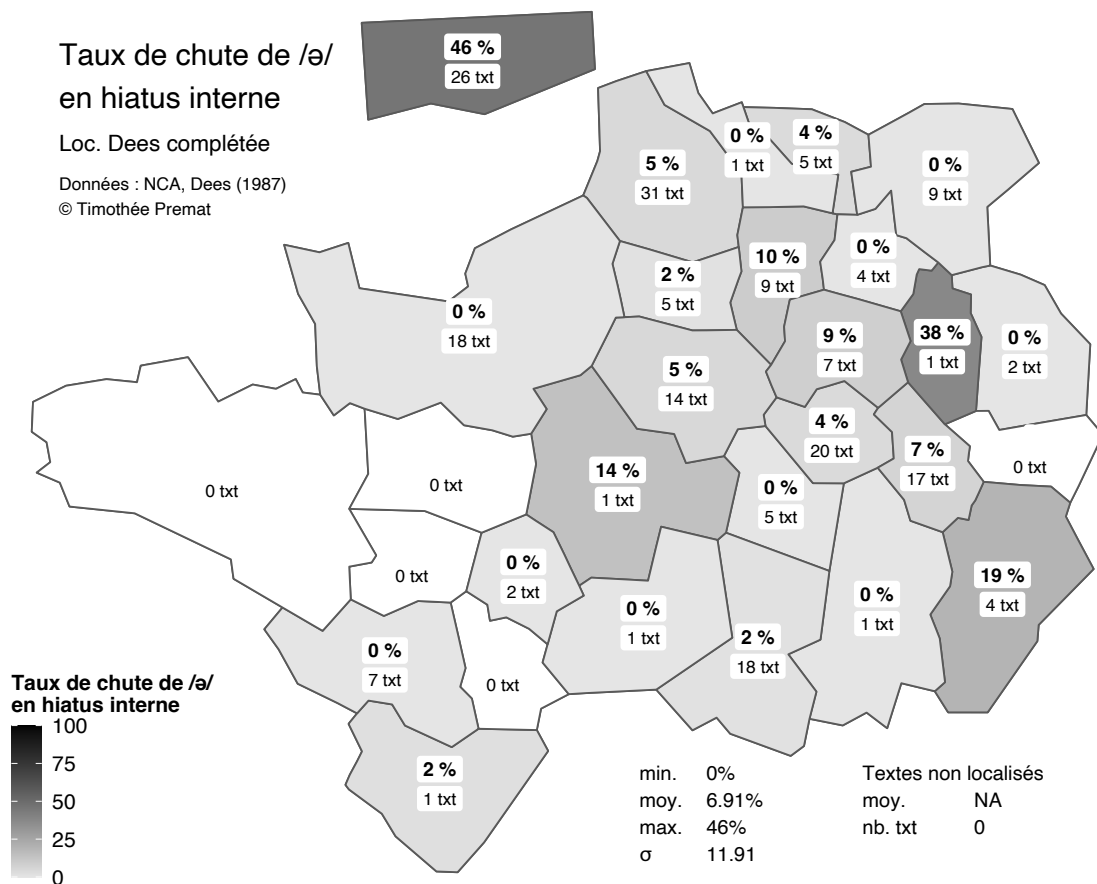


FIGURE 11.2 – Diatopie des taux de réduction des hiatus əV, sur l'ensemble de la période considérée.

La carte en fig. 11.2 représente la distribution diatopique de graphies réduites (<V> au lieu de <əV>), pour l'ensemble de la période considérée, de 1100 à 1322. Seules trois régions se détachent clairement en termes de taux de réduction des hiatus əV : l'Angleterre, la Meuse et la Franche-Comté. Ces trois régions sont les seules à avoir des taux supérieurs à 15%, seuil lui-même bien supérieur à la moyenne des taux régionaux (moy. = 6,91%), mais elles sont peu homogènes entre-elles, que ce soit en nombre de textes (Meuse : 1 texte ; Franche-Comté : 4 textes ; Angleterre : 26 textes) ou en taux de graphie réduite (Franche-Comté : 19% ; Meuse : 38% ; Angleterre : 46%).

Ainsi, dans nos données, ce n'est pas le Nord du domaine en général qui montre un taux de graphies réduites plus élevées, mais c'est surtout l'Angleterre. Les autres régions du Nord du domaine ont des taux qui, à l'exception de la Meuse, ne dépassent pas 10% et sont nuls dans plusieurs régions. Certains de ces taux nuls ou peu élevés sont appuyés par un nombre de textes assez élevé, comme par exemple la Normandie (0%, 18 textes), la Somme et le Pas-de-Calais (5%, 31 textes), ou la Wallonie (0%, 9 textes). Il faut donc soit rejeter la prééminence du Nord dans la réduction des hiatus, soit la reformuler en la faisant porter sur l'Angleterre seule ;

c'est d'ailleurs sur la singularité et la précocité de l'anglo-normand sur ce point qu'insiste SHORT (2007, p. 92-93). Le même auteur note d'ailleurs que le hiatus est si instable en anglo-normand qu'il est peut-être, plus souvent qu'en français continental, brisé par une approximante épenthétique ou un <h>, que ce soit pour des hiatus à deux voyelles pleines (*oyer* pour *oïr*, *raiendrat* pour *raendrat*, *puiur* pour *püur*), ou impliquant un schwa final (*espeye* ou *espeie* pour *espee*, *nuwe* pour *nue*; *juhe* pour *joue* (INDPRÉ3sg), y compris à la rime (*entreie* (*entree*) : *veie*, *seie*; *entree* : *donee*; SHORT 2007, p. 94; voir aussi POPE 1966 [1934], p. 452, 460, et cf. 11.5 e-h *supra*). L'anglo-normand montre donc un degré de tolérance au hiatus qui semble inférieur à celui de l'oïl continental, et peut résoudre ce hiatus soit par la syncope (cartographiée en fig. 11.2), soit par l'épenthèse consonantique.

De manière générale, les taux de graphies réduites sont organisés selon un large axe allant du Nord-Ouest (Angleterre) au Sud-Est (Franche-Comté); en marge de cet axe, aucune région ne dispose de plus de 5% de graphies réduites, et c'est aux extrémités de cet axe que se concentrent les taux les plus élevés. Au sein même de cet axe, toutes les régions ne sont pas concernées : même avec les données de localisation additionnelles, la Normandie, la Somme et le Pas-de-Calais, l'Oise, les Ardennes, la région parisienne et l'Aisne disposent de taux inférieurs à la moyenne des régions. Singulièrement, la région parisienne appartient à ces régions aux taux inférieurs à la moyenne, ce qui, combiné avec son voisinage avec des régions aux taux plus élevés (Orléanais : 14%; Aisne : 10%; Marne : 9%), indique que l'Île-de-France ne peut être considérée comme le centre de diffusion de cette réduction graphique. Nous verrons dans la section suivante que le développement de cet axe est confirmé en diachronie.

11.2.3 Intersection de la diachronie et de la diatopie

Pour observer le développement diachronique de cette répartition diatopique, nous avons arbitrairement divisé la période temporelle considérée en trois séquences : deux de 75 ans (de 1100 à 1175 et de 1175 à 1250) et une de 72 ans (de 1250 au dernier texte, datant de 1322). Étant donné le faible nombre de lemmes et d'occurrences considérées, nous avons adopté cette répartition par fenêtre de deux tiers de siècle pour ne pas appauvrir trop fortement les données de chaque carte. Ces périodes sont représentées en fig. 11.3, 11.4 et 11.5, en pages suivantes.

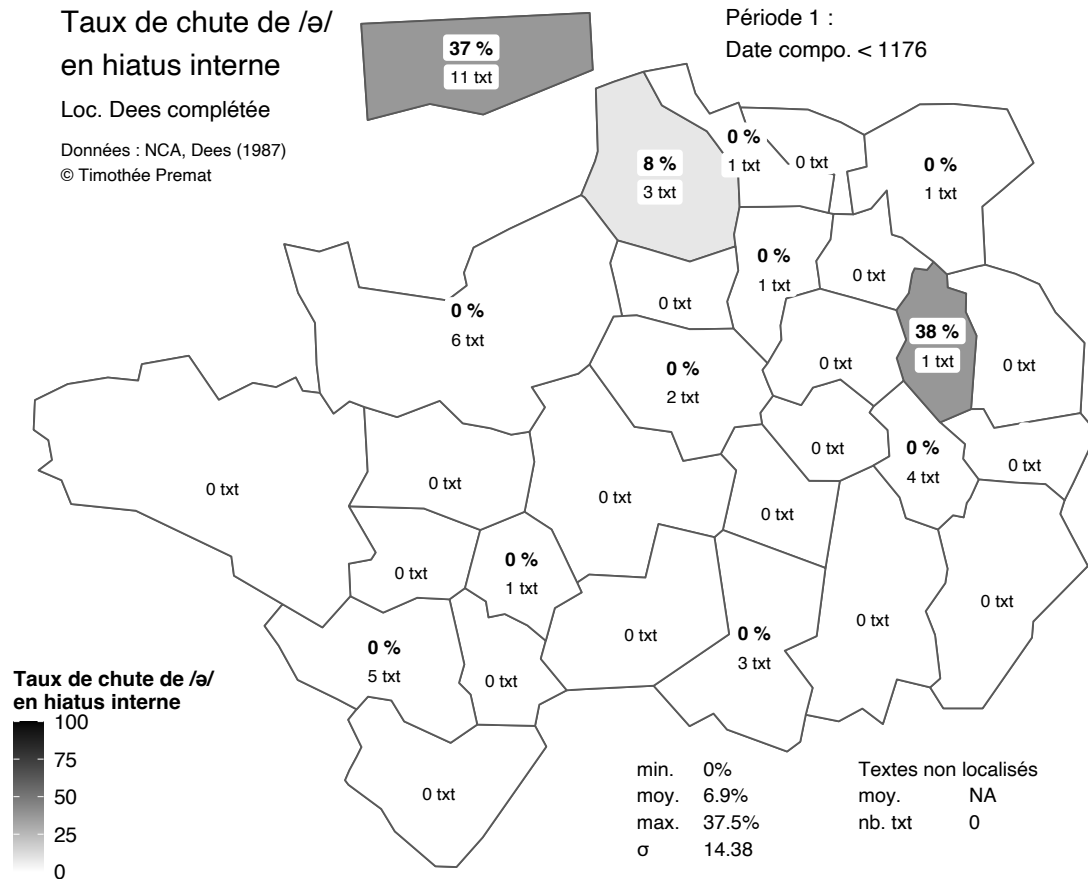


FIGURE 11.3 – Diatopie des taux de réduction des hiatus əV, période 1 (1100-1175).

11.2.3.1 1100-1175

La précocité de l'anglo-normand est confirmée par nos données : avant 1176, l'Angleterre dispose déjà d'une moyenne de 37% de graphies réduites (fig. 11.3). Ni md, le seul texte meusien répondant à nos lemmes hiatiques, est déjà présent, sa composition datant de 1150. Ce n'est pas le texte dont le taux de graphies réduites est le plus élevé ; 3 textes ne disposent que de graphies réduites (*psautier*, *deusamH*, *yoneH*), et un texte dispose de 85% de graphies réduites (*teo2*). C'est en revanche le seul texte à graphies réduites de sa région, ce qui provoque la moyenne très élevée de la Meuse (38%). Hormis ces deux régions, seule la région Somme et Pas-de-Calais dispose de formes à graphie réduite avant 1176, avec 8% de graphies réduites sur 3 textes. Cette moyenne résulte de la co-existence de *nimc* (1 occ. à graphie réduite pour 4 occ. à graphie non réduite) avec deux textes ne disposant d'aucune graphie réduite, *guigP* (3 occ. non réduites) et *aiol* (1 occ. non réduite). La fiabilité statistique de toutes ces régions, hormis l'Angleterre, est donc faible. Néanmoins, dès cette première période, se dessine une partie de l'axe Nord-Ouest - Sud-Est mentionné *supra* : le pôle anglais domine déjà (37%, 11 textes), et la région la plus proche, Somme et Pas-de-Calais, dispose déjà d'attestations de la réduction des hiatus. Dans tout le reste du domaine, la réduction est non attestée, soit parce qu'aucun texte ne répond à nos huit lemmes

deusamH : MARIE DE FRANCE (1165*). *Les Deus Amanz*. Ms. H; Dees : nil; compo. a.n., ca 1165; ms. a.n., ca 1275; MS (transcr. inédite sans transcripteur indiqué).

yoneH : MARIE DE FRANCE (1165*). *Lai de Yonec*. Ms. H; Dees : nil; compo. a.n., ca 1165; ms. a.n., ca 1275; MS (éd. non identifiée).

teo2 : ADGAR (1150*). *Le dit de Théophile*. Ms. M; Dees : nil; compo. a.n., ca 1150; ms. a.n., ca 1210; MS2 (WEBER 1877).

guigP : MARIE DE FRANCE (1165*). *Le Lai de Guigemar*. Ms. P; Dees : nil; compo. N.-O., ca 1165; ms. pic., ca 1290; nil (transcr. inédite sans transcripteur indiqué).

aiol : ANON. (1160*). *Aiol*. Dees : Somme, Pas-de-Calais; compo. pic., ca 1160; ms. pic., ca 1275; MS1 (NORMAND ET RAYNAUD 1877); DEAF : Aiol^{1/2}N.

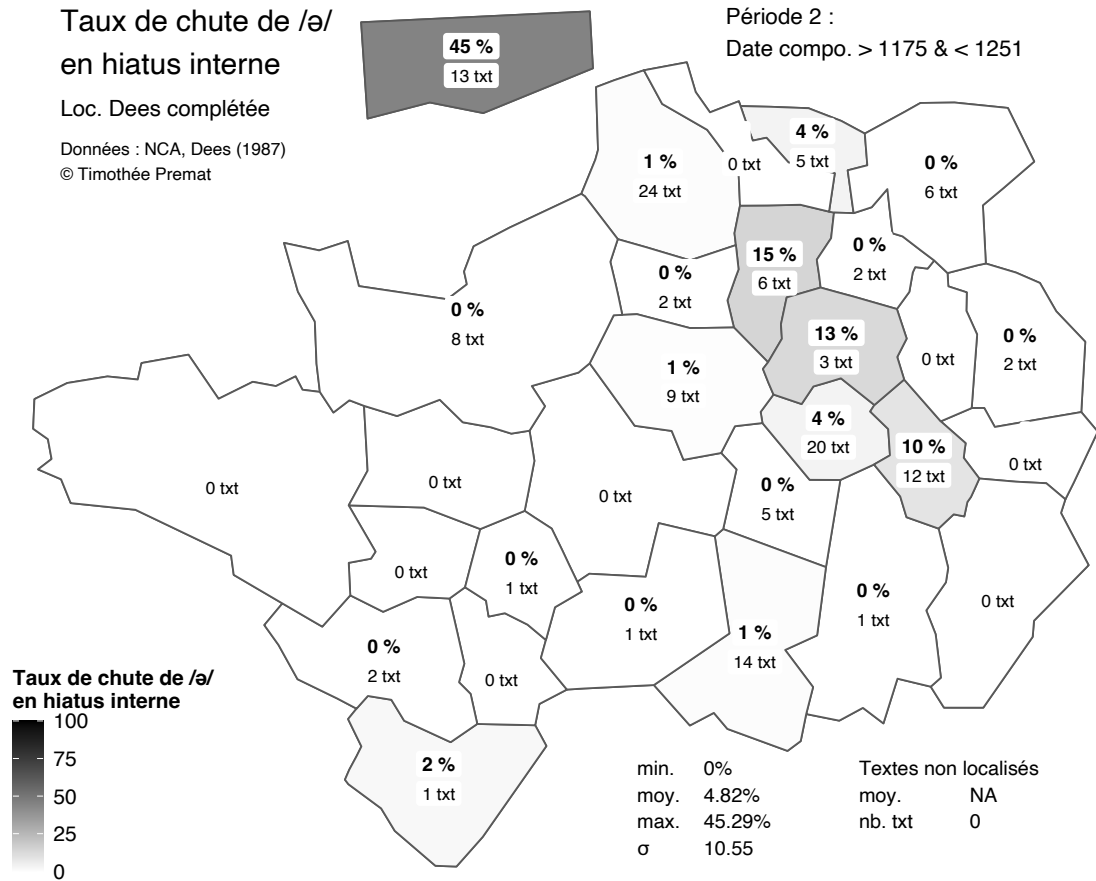


FIGURE 11.4 – Diatopie des taux de réduction des hiatus əV, période 2 (1176-1250).

hiatiques, soit parce que les textes répondant à ces lemmes ne disposent que d'occurrence non réduites (Normandie, région Somme, Pas-de-Calais, Aisne, région parisienne, Wallonie, Haute-Marne, région Nièvre, Allier, Indre-et-Loire et région Vendée, Deux-Sèvres).

11.2.3.2 1176-1250

Dans la deuxième période (fig. 11.4), la prédominance de l'Angleterre se renforce (45%, 13 textes), tandis que le tracé de l'axe Nord-Ouest - Sud-Est se précise. En effet, on ne trouve de taux supérieurs à 2% sur le continent qu'entre le Hainaut (4%, 5 textes) et la Haute-Marne (10%, 12 textes), avec l'Aisne (15%, 6 textes), la Marne (13%, 3 textes) et l'Aube (4%, 20 textes). Le nombre de textes de ces régions varie, mais certaines disposent d'une assise empirique satisfaisante (Haute-Marne, Aisne, Aube > 10 textes). Ailleurs, des taux marginaux apparaissent, que ce soit à proximité (région parisienne : 1%, 9 textes) ou à distance (Nièvre et Allier : 1%, 14 textes ; Charente-Maritime : 2%, 1 textes) de cet axe. Enfin, le taux de la région Somme et Pas-de-Calais est en nette régression, passant de 8% à 1%. Cette régression est due au fait qu'un seul texte sur 24, pon1, dispose d'une occurrence réduite, pour 3 occurrences non réduites ; la réduction du taux est donc directement liée à l'augmentation du nombre de textes.

pon1 : ANON. (1213*). *La fille du comte de Pontieu*. Première main ; Dees : Somme, Pas-de-Calais ; compo. Nord, ca 1213 ; ms. Nord, ca 1290 ; MS2 (BRUNEL 1926).

11.2.3.4 En somme

Ainsi, l'axe déjà peu homogène que nous avons observé sur les cartes synchroniques (fig. 11.2) semble se décomposer en trois mouvements. La phénomène apparaît d'abord au Nord-Ouest (Angleterre et, dans une moindre mesure, région Somme, Pas-de-Calais), accompagné d'une occurrence isolée dans la Meuse avec *nimd*. Il s'étend ensuite selon un axe Nord-Ouest - Sud-Est assez bien dessiné, mais rompu par la résurgence de formes non réduites dans la Somme et le Pas-de-Calais, tandis que des textes sporadiques à faibles taux de graphies réduites apparaissent à distance de l'axe (Charente-Maritime, Nièvre et Allier). Enfin, l'intégrité de l'axe est rompu en son centre, alors que certaines de ses marges intègrent l'innovation. Parmi elles, l'extrémité Sud-Est de l'axe (l'innovation atteint cette région alors que l'axe se désagrège), et le Sud de l'axe avec la région parisienne et l'Orléanais ; mais dans le cas de ces régions, le nombre de textes impliqués est peu satisfaisant. Durant toute cette évolution, l'Angleterre conserve un taux de formes réduites toujours largement supérieur à la moyenne et, si l'on fait exception de la Meuse en période 1 avec son attestation isolée, conserve le taux le plus élevé du domaine. D'autres régions ne sont jamais atteintes par l'innovation, qu'elles soient à proximité immédiate de l'axe (Normandie) ou non (Wallonie).

La réduction graphique des hiatus *əV* n'est donc jamais un phénomène homogène sur le plan diatopique ; l'interaction entre diachronie et diatopie est ici forte pour toutes les régions sauf pour l'Angleterre, l'axe Nord-Ouest - Sud-Est de diffusion semblant se désintégrer entre la période 2 et la période 3 et s'étendre vers le Sud en période 3. Malgré quelques attestations sporadiques, la dynamique est assez claire, et ne saurait relever d'un hasard statistique. Il s'agit donc bien, selon toute vraisemblance, d'une innovation grapho-phonologique, qui se répand sur une partie du territoire et, dans certaines régions, régresse ensuite. En ce sens, il faut non pas évacuer mais préciser la remarque diatopique de ZINK (2013 [1986], p. 164) et POPE (1966 [1934], p. 116) : la prééminence nordique de la suppression des schwas hiatiques ne concerne pas le Nord en général mais concerne surtout l'anglo-normand, et la répartition diachronique de cette suppression prend les caractéristiques d'une innovation, fortement présente en anglo-normand et plus discrètement dans d'autres dialectes septentrionaux, qui se répand selon un axe Nord-Ouest - Sud-Est, ignorant notamment le Nord-Est du domaine.

11.3 Chute de schwa internes hors hiatus

La chute de schwa interne non hiatique est attestée par de nombreuses grammaires, mais, comme exposé en sec. 11.1.1.2 *supra*, il y a une divergence entre les auteurs pour l'analyse des contextes qui conditionnent cette syncope. Si l'on laisse de côté la question de la définition des contextes déclencheurs, la divergence sur la datation de la syncope est en revanche moins prononcée que pour les schwas internes en hiatus, mais cela est notamment dû au fait que peu de datations sont proposées. Par ailleurs, puisque les auteurs ne sont pas d'accord sur les détails des contextes qui déclenchent la syncope, ils ne peuvent pas non plus

être d'accord sur la datation de l'application de la syncope à chaque contexte, à moins de s'en tenir à noter que la syncope se développe « au cours du Moyen Âge » comme le fait NYROP (1899, p. 212-213). Sur le plan diatopique, peu d'indications nous sont données, Zink mentionnant une syncope aux contextes moins restrictifs en picard du XIII^e s.; par ailleurs, ici, la spécificité de l'anglo-normand ne peut s'analyser en termes de datation dialectale, puisque si les premières attestations sont anglo-normandes, elles datent du XI^e s., c'est-à-dire à une époque où la production continentale n'en est qu'à ses balbutiements.

Nous adoptons ici la même méthode que pour la section précédente : après avoir relevé ce que la bibliographie dit de la diachronie et de la diatopie de la syncope, nous effectuons un sondage sur huit lemmes et observons la distribution diachronique et diatopique de la syncope dans ces lemmes. Cette section commence donc par explorer la diachronie des syncopes (sec. 11.3.1), en développant la méthode utilisée (sec. 11.3.1.1) avant d'aborder les premières attestations syncopées (sec. 11.3.1.2) et le développement diachronique de la syncope sous la forme de la co-existence entre formes syncopées et non syncopées (sec. 11.3.1.3). Cette section aborde ensuite la diatopie de la syncope (sec. 11.3.2) et l'intersection entre diatopie et diachronie (sec. 11.3.3).

11.3.1 Diachronie de la réduction des schwas hors hiatus

La datation de la chute de schwas internes interconsonantiques, c'est-à-dire leur syncope, n'est guère plus consensuelle que celle des chutes de schwas hiatiques, notamment à cause de la pluralité des contextes impliqués et de l'absence de consensus sur la définition et le rôle de chaque contexte (cf. sec. 11.1.1.2). Néanmoins, les auteurs qui s'expriment à propos de la datation du phénomène semblent d'accord sur un point : la syncope ne concerne, au Moyen Âge, qu'un ensemble restreint de contextes, et il faut attendre la fin du Moyen Âge pour qu'elle se libère de ces contextes. Ainsi, selon FOUCHÉ (1969 [1958], p. 524), il faut attendre le XV^e s. pour que la syncope s'applique sans que son environnement ne contienne de liquide. Sauf à considérer que les nasales sont des liquides, cela est en contradiction avec une partie des descriptions : POPE (1966 [1934], p. 118) donne les débuts des syncopes dans des environnements contenant {/n, r, l/} à la fin du XII^e et au XIII^e s. ZINK (2013 [1986], p. 189) et SHORT (2007, p. 95) font remonter le *terminus a quo* de la syncope au XI^e s. avec, pour Short, le vers d'*Alexis* cité en (11.11 a) *supra*. Par ailleurs, rejoignant ici la dynamique que décrit Fouché, Zink indique que les contextes déclencheurs de la syncope s'élargissent, pour concerner, en picard du XIII^e s., les contextes contenant un /r/, quelle que soit l'autre consonne impliquée. Il faut donc s'attendre à trouver les premières attestations de la syncope au XI^e s. et, éventuellement, à ce que les premières attestations soient décalées dans le temps en fonction des contextes qui déclenchent la syncope. Par ailleurs, comme évoqué en sec. 11.1.1.2, la syncope peut être graphique et métrique (c'est-à-dire, selon toute vraisemblance, graphique et phonologique; cf. 11.11 a *supra*), ou n'être suggérée que par la métrique, la forme graphique conservant un <e> qui n'a pas de valeur métrique (11.11 a *supra* : *trovereiz* bisyllabique).

11.3.1.1 Méthode

Comme pour les schwas internes hiatiques, nous avons, face à l'absence de doctrine claire, recherché dans le NCA les occurrences syncopées de lemmes contenant un schwa interne interconsonantique. Une nouvelle fois, nous avons sélectionné huit lemmes ou familles de lemmes. Comme illustré en (11.13), il ne s'agit pas à strictement parler de huit lemmes, dans la mesure où des lemmes dérivés par adjonction suffixale sont également pris en compte (11.13 h). Lorsqu'il y a une variation graphique parmi les formes relevées, l'ensemble des formes considérées est donné en note. L'annotation en temps et modes indiquée dans les notes est celle de Dees, nous ne l'avons pas vérifiée dans la mesure où les contextes phonologiques sont constants. Nous avons exclu toutes les formes dont l'annotation morphosyntaxique et/ou en lemme était discordante ; ainsi, les occurrences *verai* dont le lemme est *voir* et la pos VER_futu_1_sg sont disqualifiées pour appartenir aux formes adjectivales *verai(es)* qui s'opposent aux formes adjectivales *vrai(es)*, pouvant représenter un IND FUT du verbe *voir*. Il en va de même pour certaines formes *durement* (lorsque lemme = *durement* mais pos = NOM_obj_mas_c_sg), disqualifiées pour appartenir aux formes adverbiales du type *durement* qui s'opposent aux formes adverbiales du type *durment*. Une exception est faite pour les lemmes à valeur unknown, lorsque l'annotation morphosyntaxique est correcte (p.ex. *vraiment* avec lemme = UNKNOWN mais pos = adv est considéré comme qualifié pour appartenir au type syncopé du lemme adverbial *vraiment* ; ces cas sont marginaux).²⁴

(11.13) Formes non hiatiques retenues

(a) <i>denier+ée</i>	→ <i>deneree</i>	→ <i>denree</i> ²⁵
(b) <i>dur+e+ment</i>	→ <i>durement</i> ²⁶	→ <i>durment</i> ²⁷
(c) <i>sporo</i> († frq., b.lat.)	→ <i>esperons</i> ²⁸	→ <i>espron</i> ²⁹
(d) <i>fol+e+ment</i>	→ <i>folement</i> ³⁰	→ <i>folment</i> ³¹
(e) <i>porter</i> + suff. IND FUT	→ <i>portera</i> ³²	→ <i>portra</i> ³³
+ suff. COND PRÉ	→ <i>portereit</i> ³⁴	
(f) SACRĀMĒNTUM	→ <i>sairement</i> ³⁵	→ <i>sairment</i> ³⁶
(g) VERTĪBŪLUM	→ <i>vertevelle</i>	→ <i>verveles</i> ³⁷
(h) <i>veracus</i> (lat.pop.)	→ <i>verai</i> ³⁸	→ <i>vrai</i> ³⁹
<i>verai+e+ment</i>	→ <i>vraiment</i> ⁴⁰	→ <i>vraiment</i> ⁴¹

Les contextes de déclenchement de la syncope représentés par les formes en (11.13) sont donnés en (11.14). Ils représentent l'essentiel des différents contextes relevés dans les grammaires et donnés précédemment (sec. 11.1.1.2), à ceci près que les liquides et nasales sont toujours associées entre elles ou avec une obstruante en précession (*contra* NYROP 1899, p. 213 ; POPE 1966 [1934], p. 118 ; GESS, JACOBS et LAKS 2020, GGHF, p. 456-478 qui ne précisent pas de contrainte sur le statut de l'autre consonne impliquée), mais en accord avec la domination des configurations OBS+LIQ relevée par certaines grammaires (SÉGÉRAL et SCHEER 2020, GGHF, p. 344 ; GESS, JACOBS et LAKS 2020, GGHF, p. 478). La sélection des huit lemmes ou familles de lemmes en (11.13) ne signifie pas que les autres contextes ne sont pas attestés ; simplement, dans notre corpus, ils ne donnaient que de trop faibles nombres d'occurrences (souvent une ou deux au plus) pour être pris en considération dans une optique statistique. Nous ne sommes pas en mesure, ici, d'évaluer ce qui, dans cette rareté, relève

24: La valeur UNKNOWN pour l'attribut lemme est déclenchée lors que TreeTagger n'a pas su associer de lemme à la forme en question, soit à cause d'une variation graphique à laquelle il n'était pas préparé, soit à cause d'un contexte trop rare, voire de la conjonction des deux configurations.

25: NOMsg *denree*, *donree* ; pl *danrees*, *denrees*.

26: *durament*, *durement*, *durement*, *durement*.

27: *durement*.

28: *esperons*, *esperrons*, *esperons*, *esperuns*.

29: NOMsg *espron* ; pl *esprons*.

30: ADV *folemant*, *folement*, *follemant*, *follement*, *foulemant*.

31: ADV *folment*.

32: IND FUT 1sg *porterei*, *porteraï* ; 2sg *porteras* ; 3sg *portera*, *porteraï* ; 1pl *porterons*, *porterum* ; 2pl *portereiz*, *portérés*, *porterez*, *porteroiz* ; 3pl *porteront*, *porterunt*.

33: IND FUT 3sg *portra*.

34: COND PRÉ 1g *porteroie* ; 3sg *portereit*, *porteroit* ; 3pl *porteroient*.

35: NOM (« serment ») *sairemant*, *sairement*, *sairementz*, *serementz*, *serement*.

36: NOM *serment*.

37: DMF : « Anneau qui tient le verrou d'une porte », etc.

38: ADJ MASCsg *verai*, *verais*, *veraix*, *verai*z ; FÉMSG *veraie*, *veraiee* ; MASCpl *verais* ; FÉMpl *verai*es, *verais*.

39: ADJ MASCsg *vrai*, *vrais* ; FÉMSG *vraie* ; MASCpl *vrai*, *vrais* ; FÉMpl *vraie*es, *vrais*.

40: ADV *vraiment*, *vraiment*, *vraiment*.

41: ADV *vraiment*, *vraiment*.

du hasard des fréquences lexicales, ou au contraire relève d'une affinité phonologique.

(11.14) Contextes de syncope des formes non hiatiques retenues

(a) Contextes OBS+LIQ

- a) /p__r/ *esperons ~ espron*
- b) /rt__r/ *portera ~ portra*
- c) /rt__v/ *vertevelle ~ verveles*
- d) /v__r/ *verai(-) ~ vrai(-)*

(b) Contextes hors OBS+LIQ

- a) /n__r/ *deneree ~ denree*
- b) /r__m/ *durement ~ durment*
sairement ~ sairement
- c) /l__m/ *folement ~ folment*

Par ailleurs, ces formes présentent des statuts différents vis-à-vis de la syncope. Dans tous les cas sauf un, la syncope n'atteint que le schwa ; la forme syncopée de *vertevelle* (11.13 g) connaît également une simplification consonantique, par chute du /t/ dans *verveles*, le groupe consonantique /rtv/ représentant un monstre phonotactique. En revanche, l'attestation de *portra* (11.13 e) indique que le groupe /rtr/ a pu être licite ici, mais cette attestation ne repose, dans notre corpus, que sur une occurrence, tandis que le corpus comporte 191 occurrences non syncopées à l'indicatif futur ou au conditionnel présent. Puisque le groupe /rtr/ est licite en a.fr., comme en attestent de nombreux lemmes (*martre, chartre, portraiture, etc.*), le caractère fort peu productif de la syncope dans *portra* doit s'expliquer par une prohibition ou restriction à la dérivation synchronique : le groupe /rtr/ est licite lorsqu'il est étymologique mais, en a.fr., il ne semble guère pouvoir être créé par une syncope synchronique.

Enfin, il faut noter que le schwa sujet à syncope peut être issu de la dérivation, soit depuis une voyelle tonique par l'adjonction à l'infinitif (qui sert alors de base) d'un suffixe tonique, soit dans la formation des adverbes pour *durement* (11.13 b) et *folement* (11.13 d), mais pas pour *véraiment* (11.13 h), où le schwa de composition adverbiale est en hiatus et où c'est le schwa radical qui est sujet à la syncope. Dans ce dernier cas, il n'est pas possible d'exclure formellement que des analogies ne jouent, dans la mesure où, lorsque l'adjectif *vérai* s'est syncopé en *vrai*, l'adverbe *véraiment* pourrait très bien passer à *vraiment* conséquemment. Comme la réciproque ne saurait être catégoriquement exclue non plus (la syncope de l'adverbe pourrait entraîner et/ou encourager la syncope de l'adjectif), nous avons choisi de traiter de l'adverbe et de l'adjectif comme s'il s'agissait d'un seul et même item.

11.3.1.2 Premières attestations de la syncope

Dans notre corpus, les premières attestations syncopées de ces huit lemmes ou bases apparaissent dans la première moitié du XII^e s., avec 5 occurrences *vrai(es)* pour 2 occurrences *vérai* dans *marga*. Ce texte, dont la composition est datée de ca 1135, est donné comme provenant de la Haute-Marne par Dees, tandis que les localisations philologiques donnent la Normandie pour la composition et l'Artois pour la copie.

marga : WACE (1135*). *La vie de Sainte Marguerite de Wace*. Ms. M; Dees : Haute-Marne; compo. norm., ca 1135; ms. art., 1267; CR3 (WACE 1932a); DEAF : WaceMargaF.

Suivent les premières attestations de *denree*, dans plusieurs textes du *Charroi de Nîmes* (cycle épique de Guillaume d'Orange) : *nima1*, *nima2* et *nimc* contiennent chacun deux occurrences *denree*; *nima3*, *nima4*, *nimb1* et *nimb2* contiennent chacun une occurrence *denree*. Les textes *nima3*, *nima4* et *nimc* comportent également une occurrence *denree* chacun. Par ailleurs, *nimb1* et *nimb2* contiennent également une forme *vraiment*, alors qu'aucun de ces textes ne contient de formes non syncopées de *verai* ou *verai(e)ment*. La présence de ces textes est d'autant plus saisissante que *nimc* et *nimd* présentaient déjà les premières attestations de la réduction du hiatus dans *march(e)ant*; il s'agit donc d'une famille de textes novateurs sur la question des suppressions de schwas prétoniques. L'homogénéité entre ces versions du *Charroi de Nîmes* vient confirmer la pertinence des dates de composition vis-à-vis de celle des dates de copie, ces dernières s'étendant de 1263 à 1335. En revanche, les localisations de Dees ne montrent pas un haut degré de cohérence entre ces textes, dans la Haute-Marne (*nima1*), la région Nièvre et Allier (*nima2*, *nima3*, *nima4*), l'Aisne (*nimb1*), la Normandie (*nimb2*) ou la région Somme et Pas-de-Calais (*nimc*). Aucune localisation philologique de composition n'étant proposée par les métadonnées du NCA pour ces textes, nous ne pouvons comparer les localisations de Dees aux localisations traditionnelles. De cette multiplicité des régions de Dees peut découler l'hypothèse que le phénomène étudié appartient en propre à la strate linguistique la plus profonde de ces textes (la strate de composition), et ne réagit pas aux strates linguistiques plus superficielles, lesquelles composent le marquage dialectal auquel se réfère Dees. Si l'on admet, en effet, que la syncope appartient à la version originelle du texte, on peut comprendre qu'elle irradie dans l'ensemble des versions depuis le haut du stemma, indifféremment au marquage dialectal qu'adopte chaque version.

La version syncopée de *sair(e)ment* apparaît dans notre corpus 15 ans après les formes syncopées de *v(e)rai*, avec une occurrence *serment* dans la version du ms. H du *Lai de Lanval* de Marie de France (*lanvalH*), dont la composition est datée de *ca* 1165. Cette unique occurrence syncopée s'oppose à une autre occurrence non syncopée *serement* dans le même texte; pour ce mot, les deux autres versions de notre corpus ne comportent que la forme non syncopée, avec deux formes *sairement* dans le ms. P (*lanvalP*) et deux formes *serement* dans le ms. S (*lanvalS*). Par ailleurs, les ms. C (*lanvalC*), H, P et S comportent également une occurrence non syncopée d'un autre lemme, *durement*. Ici, et contrairement à ce qui se passe dans la famille des ms. du *Charroi de Nîmes*, la syncope de *sair(e)ment* ne peut être considérée comme une innovation du nœud le plus élevé du stemma qui se serait maintenue dans les nœuds inférieurs, dans la mesure où elle n'est représentée que par un seul témoin. Par ailleurs, la diffusion lexicale est ici en cours: alors que *sair(e)ment* et *dur(e)ment* disposent du même contexte /Vr__m/, l'un est déjà atteint par la syncope et l'autre pas.

Alors que les formes syncopées des lemmes *verai-*, *denree* et *sairement* réapparaissent ensuite régulièrement dans la seconde moitié du XII^e s., il faut attendre le début du XIII^e s. pour qu'apparaisse notre unique forme *portra*, dans *moral*, texte wallon (ici, en consensus entre Dees et les deux localisations philologiques) dont la composition est datée de *ca* 1200 et la copie de *ca* 1210. Ce texte contient également 31 occurrences syncopées de la base *v(e)rai*, et seulement deux occurrences non syncopées de nos 8

nima1 : ANON. (1150*). *Le charroi de Nîmes*. Ms. A1; Dees : Haute-Marne; compo. nil, *ca* 1150; ms. frc., *ca* 1263; MS1 (transcr. inédite L. Schoesler).

nima2 : ANON. (1150*). *Le charroi de Nîmes*. Ms. A2; Dees : Nièvre, Allier; compo. nil, *ca* 1150; ms. frc., *ca* 1263; MS1 (transcr. inédite L. Schoesler).

nimc : ANON. (1150*). *Le charroi de Nîmes*. Ms. C; Dees : Somme, Pas-de-Calais; compo. nil, *ca* 1150; ms. art., 1295; MS1 (transcr. inédite L. Schoesler).

nima3 : ANON. (1150*). *Le charroi de Nîmes*. Ms. A3; Dees : Nièvre, Allier; compo. nil, *ca* 1150; ms. lorr., *ca* 1325; MS1 (transcr. inédite L. Schoesler).

nima4 : ANON. (1150*). *Le charroi de Nîmes*. Ms. A4; Dees : Nièvre, Allier; compo. nil, *ca* 1150; ms. frc., *ca* 1283; MS1 (transcr. inédite L. Schoesler).

nimb1 : ANON. (1150*). *Le charroi de Nîmes*. Ms. B1; Dees : Aisne; compo. nil, *ca* 1150; ms. Paris, *ca* 1335; MS1 (transcr. inédite L. Schoesler).

nimb2 : ANON. (1150*). *Le charroi de Nîmes*. Ms. B2; Dees : Normandie; compo. nil, *ca* 1150; ms. Paris, *ca* 1335; MS1 (transcr. inédite L. Schoesler).

lanvalH : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. H; Dees : nil; compo. N.-O., *ca* 1165; ms. a.n., *ca* 1275; MS2 (MARIE DE FRANCE 1958); DEAF : *MarieLanvR*².

lanvalP : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. P; Dees : nil; compo. N.-O., *ca* 1165; ms. pic., *ca* 1290; MS2 (MARIE DE FRANCE 1958); DEAF : *MarieLanvR*².

lanvalS : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. S; Dees : nil; compo. N.-O., *ca* 1165; ms. frc., *ca* 1300; MS2 (MARIE DE FRANCE 1958); DEAF : *MarieLanvR*².

lanvalC : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. C; Dees : Angleterre; compo. N.-O., *ca* 1165; ms. a.n., *ca* 1290; MS2 (MARIE DE FRANCE 1958); DEAF : *MarieLanvR*².

moral : ANON. (1200*). *Le poème moral, traité de la Vie Chrétienne*. V. 1-2320; Dees : Wallonie; compo. wall., *ca* 1200; ms. wall., *ca* 1210; CR1 (BAYOT 1929); DEAF : *PoèmeMorB*.

lemmes : *folment* et *durement*. Il est donc très productif sur le plan de la syncope, avec 94% de formes de nos huit lemmes qui y sont syncopées. Il semble pertinent de noter que la syncope exceptionnelle dans *port(e)ra*, acceptable au plan phonotactique mais très marquée et non attestée dans le reste du corpus, a lieu dans l'un des textes les plus productifs sur le plan de la syncope, et dans lequel seuls des lemmes qui n'ont pas encore de formes syncopées par ailleurs peuvent préserver leur schwa interne non hiatique.

À la même date de composition (ca 1200), mais avec plus d'un siècle d'écart pour la date de copie (ca 1370), apparaît la première forme syncopée *durment*, dans un texte du Pas-de-Calais, *or*. Ce texte est également fort productif sur le plan de la syncope, avec cinq formes syncopées de la base *v(e)rai* et une seule forme non syncopée de nos huit lemmes, *portera*. Les premières occurrences syncopées de *ver(te)v(e)lle* apparaissent une quinzaine d'années plus tard, avec deux occurrences *verveles* dans *clari2*, texte picard (Dees : Somme et Pas-de-Calais), et dont la composition est datée de ca 1217 et la copie de ca 1300. Ce texte ne contient qu'une seule autre occurrence syncopée de nos huit lemmes (*vraie*), et comporte 23 occurrences non syncopées *durement* et 3 occurrences non syncopées *esperons*; la syncope de *durement* n'est attestée dans notre corpus que peu avant mais dans la même région, tandis que *espron* sera attesté peu après dans la même région également. Ces deux occurrences de *verveles* sont difficiles à analyser, étant donné que le lemme est très rare dans notre corpus : aux deux occurrences de *clari2* s'ajoute une seule occurrence, non syncopée, dans un texte postérieur (*ndchar*).

Les premières occurrences syncopées du lemme *esp(e)ron* apparaissent peu après, avec *desp*, texte picard dont la composition est datée de ca 1238 et la copie de ca 1300. Ce texte ne contient aucune autre forme syncopée, mais contient 13 formes non syncopées *durement*, ainsi que deux formes *porterai* et *porteroie*. Enfin, les formes syncopées de *fol(e)ment* sont restreintes à un texte de la fin du XIII^e s., *wallo*, qui contient deux occurrences *folment* et aucune autre occurrence de nos huit lemmes, qu'elles soient syncopées ou non. Ce texte est daté, pour la composition comme la copie, de ca 1290, et est localisé, en consensus entre Dees et les deux localisations philologiques, en Wallonie. L'ensemble de ces premières attestations est résumé en table 11.2.

TABLE 11.2 – Premières occurrences de formes syncopées dans le NCA.

Lemme	Première occurrence			
	Date compo.	Date ms.	Région	ID
<i>v(e)rai(-)</i>	1135	1267	Haute-Marne	marga
<i>den(e)ree</i>	1150	≥1263	multiples	nimal-4, -b1-2,-c
<i>sair(e)ment</i>	1165	1275	Angleterre	lanvalH
<i>port(e)r-</i>	1200	1210	Wallonie	moral
<i>dur(e)ment</i>	1200	1310	Pas-de-Calais	or
<i>ver(te)v(e)lle</i>	1217	1300	Somme, P.-d.-C.	clari2
<i>esp(e)ron</i>	1238	1300	Pas-de-Calais	desp
<i>fol(e)ment</i>	1290	1290	Wallonie	wallo

Ainsi, au sein de ce panorama des premières occurrences, trois lemmes se détachent clairement : la base *porter-*, avec une seule occurrence

or : RENAULT (1200*). *La vie de saint Jehan Bouche d'Or*. Dees : Pas-de-Calais; compo. pic., ca 1200; ms. pic., 1370; CR1 (RENAULT 1931); DEAF : *SJeanBoucheD*.

clari2 : ROBERT DE CLARI (1217*). *La conquête de Constantinople*. Dees : Somme, Pas-de-Calais; compo. pic., ca 1217; ms. pic., 1300; MS (ROBERT DE CLARI 1956); DEAF : *RobClariL*.

ndchar : JEAN LE MARCHANZ (1157*). *Miracles de Notre Dame de Chartres*. Dees : Orléanais; compo. Chartres, ca 1157; ms. Chartres, ca 1350; MS2 (JEAN LE MARCHANZ 1973); DEAF : *MirNDChartreK*.

desp : ANON. (1238*). *Li chevaliers as deus espees*. Dees : Pas-de-Calais; compo. pic., ca 1238; ms. Pas-de-Calais, ca 1300; CR2 (FOERSTER 1877); DEAF : *CheviiEspF*.

wallo : ANON. (1290*). *Gloses wallonnes*. Dees : Wallonie; compo. wall., ca 1290; ms. wall., ca 1290; MS (WILMOTTE 1891).

syncopée dans un seul texte, le lemme *ver(te)velle*, avec deux occurrences syncopées et une seule occurrence non syncopée, et le lemme *fol(e)ment*, avec seulement deux occurrences syncopées dans un seul texte. Le statut de la source de ces formes ne peut guère être comparé : alors que *portra* apparaît dans un texte où la syncope est très active, *verveles* et *folment* apparaissent dans des textes répondant mal à nos interrogations, où ni la syncope ni l'absence de syncope ne sont réellement attestées pour nos autres lemmes. Enfin, *portra* et *verveles* se situent au milieu de la période des premières attestations, tandis que *folment* est tardif, à la fin du XIII^e s. ; la première attestation de *folment* est séparée des autres premières attestations par un écart de 52 ans, alors qu'entre les autres attestations, l'écart maximal entre dates voisines est de 35 ans.

Le panorama général que l'on peut dégager à partir de ces premières attestations est le suivant : d'une part, les premières attestations datent dans notre corpus de la première moitié du XII^e s., ce qui n'est guère surprenant étant donné que la bibliographie relève d'autres attestations dès le très ancien français, dans *Alexis*. D'autre part, les lemmes dont la syncope est la plus productive sont ceux dont elle est aussi la plus précoce : les formes syncopées de la base *v(e)rai-*, que l'on retrouve en ample quantité sur l'ensemble de la période considérée, sont les premières à apparaître (pour cette base : au total, 301 occurrences syncopées pour 288 occurrences non syncopées). Cela vaut également, et à plus forte raison, pour la syncope de *den(e)ree*, au nombre de 34 dans notre corpus, contre 3 formes non syncopées. Ce paradigme est rompu dans la seconde moitié du XII^e s., avec la syncope de *sair(e)ment*, qui ne concerne que 34 formes, contre 309 formes non syncopées. La syncope des lemmes dont les premières occurrences syncopées sont postérieures demeure marginale, avec 10 formes syncopées contre 688 formes non syncopées pour *dur(e)ment*, 2 occurrences syncopées contre 166 pour *esp(e)ron*, et des attestations trop ponctuelles pour les deux types pour *ver(te)velle* et *fol(e)ment*. Hormis pour *den(e)ree* dans les versions du *Charroi de Nîmes*, on note que toutes les régions où apparaissent ces premières occurrences appartiennent au Nord du domaine d'oïl, à condition de retenir la localisation philologique normande pour *marga*. Enfin, du point de vue des contextes phonologiques, on ne note pas d'effet spécifique selon que le contexte comporte un groupe OBS+LIQ ou non ; la diffusion lexicale de la syncope ne semble donc pas se faire à des vitesses différentes selon le contexte, pour peu que celui-ci dispose d'une réalisation bien formée sur le plan phonotactique.

11.3.1.3 Cooccurrences des graphies syncopées et non syncopées

Sur le plan de la co-existence entre graphies syncopées et non syncopées, la situation est quelque peu analogue à celle des suppressions de schwas internes en hiatus, c'est-à-dire que l'apparition des formes syncopées ne parvient pas à évincer les formes non syncopées, et celles-ci co-existent, que ce soit au sein du même texte (comme illustré pour certaines des premières attestations), ou au sein de différents textes à même période. Ainsi, conformément à ce que suggèrent ou disent explicitement NYROP (1899, p. 213), POPE (1966 [1934], p. 118), FOUCHÉ (1969 [1958], p. 514-515) et ZINK (2013 [1986], p. 189), on trouve une forte co-existence entre les deux types de formes à partir des premières attestations, et, selon Fouché,

jusqu'aux ^{xvi}e et ^{xvii}e s. Sur l'ensemble de la période considérée et sur l'ensemble de nos huit lemmes ou familles de lemmes, on trouve 1722 formes non syncopées pour 361 formes syncopées, mais ce taux est, comme abordé précédemment, fortement variable selon le lemme en question, de 51% de formes syncopées pour *v(e)rai* et *v(e)rai(e)ment* (587 occurrences totales) à près de 0% de formes syncopées pour *port(e)r-* (1 seule forme syncopée, *portra*, pour 189 formes non syncopées). La moyenne des taux de syncope entre les lemmes et entre les textes est supérieure à celle des taux de réduction des hiatus internes, s'établissant ici à 17% de formes syncopées (contre 9% de formes à hiatus réduits), mais ces chiffres ne sauraient être représentatifs par eux-mêmes, étant issus d'un échantillonnage. En revanche, on peut considérer que c'est l'évolution des taux de syncope qui est, ici, représentative. Celle-ci est figurée par la régression locale en fig. 11.6.

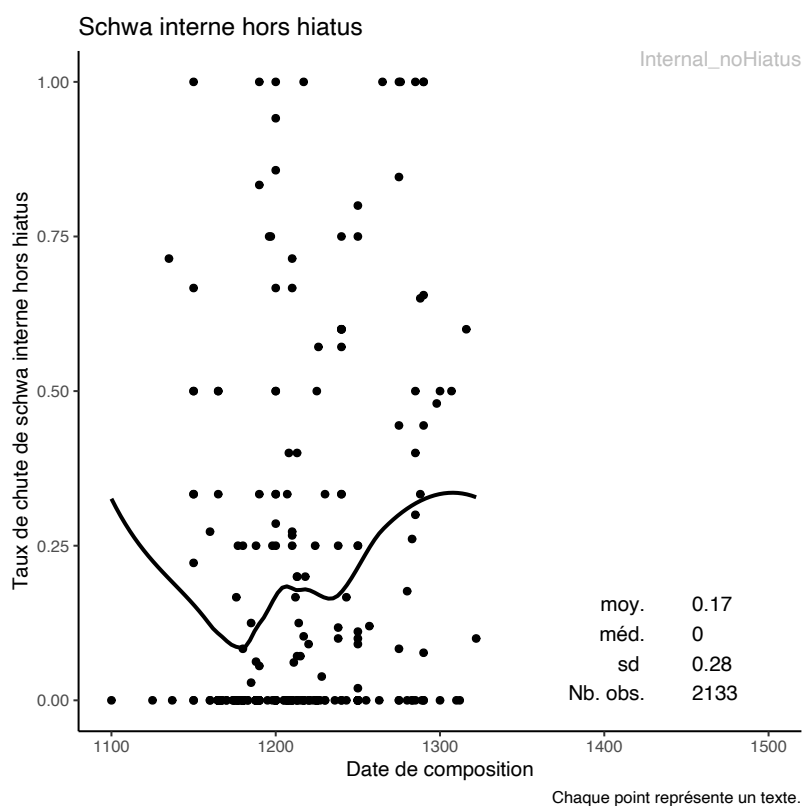


FIGURE 11.6 – Distribution diachronique des occurrences de hiatus réduits dans le NCA.

Sur la figure 11.6, il importe de considérer, comme pour la réduction des hiatus internes, que le premier mouvement de chute de la courbe de régression est principalement dû à la fenêtre sur laquelle la régression s'applique : les taux de syncope sont nuls avant 1135, et les deux textes qui précèdent (*rolandox* en 1100 et *psautier* en 1125) ne disposent d'aucune syncope sur nos 8 lemmes. Ce mouvement descendant de la régression locale est donc dû à la fenêtre mobile sur laquelle la courbe de régression se dessine, et n'est pas, ici, pertinent. En revanche, à partir du premier creux de la courbe, à la fin du ^{xii}e s., la régression dessine une courbe en croissance presque continue, certes non linéaire, mais qui nous semble représentative de la situation qu'elle figure : les taux de formes syncopées croissent, passant des alentours de 10% en moyenne à près de 30%.

rolandox : ANON. (1100*). *La Chanson de Roland*. Ms. O; Dees : Angleterre; compo. N.-O., ca 1100; ms. a.n., ca 1137; CR1 (SEGRE 1971); DEAF : Ro1S.

psautier : ANON. (1125*). *Livre des psaumes*. Dees nil; compo. a.n., ca 1125; ms. a.n., ca 1150; CR2 (MICHEL 1876), DEAF : PsCambrM.

On observe donc, ici, une évolution relativement similaire à celle des taux de réduction des hiatus internes : dans les deux cas, les schwas tendant à disparaître de plus en plus, de 10% à la fin du ^{xii} s. pour les schwas non hiatiques et au début du ^{xiii} s. pour les schwas hiatiques, aux alentours de 30% pour les schwas non hiatiques et aux alentours de 20 ou 25% pour les schwas hiatiques dans la première moitié du ^{xiv} s. C'est peut-être cette croissance que décrit la remarque de NYROP (1899, p. 212-213) selon lequel la syncope se développe « au cours du Moyen Âge », à condition de considérer qu'il ne fait pas simplement référence à l'apparition du phénomène mais au développement de son ampleur.

Également comme pour les schwas hiatiques, on retrouve ici une distribution singulière des *outliers*, c'est-à-dire des textes qui se concentrent aux extrémités de la zone de variation, avec 157 textes, répartis sur toute la période, ne témoignant d'aucune syncope (alors qu'ils témoignent de formes non syncopées), et 10 textes, à partir du milieu du ^{xii} s., ne témoignant que de formes syncopées. Dans les deux cas, cette présence d'*outliers* à 100% de schwas supprimés peut s'analyser comme participant de la tendance croissante des taux de suppression du schwa : plus le taux global d'omission du schwa est fort, plus on va trouver de textes qui l'omettent dans tous les cas. En revanche, la réciproque n'est pas nécessairement vraie, dans la mesure où, comme pour les schwas internes hiatiques, les textes à 0% de formes syncopées continuent d'exister alors que la moyenne croît, et leur raréfaction peut n'être qu'un effet de fréquence lexicale et/ou de distribution du corpus, celui-ci comportant moins de textes à son terme qu'en son centre. Du reste, les textes à 0% de formes syncopées sont 15 fois plus nombreux que les textes à 100% de formes syncopées, selon un rapport analogue à celui des taux de réduction des hiatus internes (173 textes sans hiatus réduit contre 9 textes n'ayant que des hiatus réduits). On peut donc en déduire qu'il existe une forte ressemblance entre la suppression des schwas internes en hiatus et celle des schwas prétoniques interconsonantiques, bien que la tendance à la croissance de la suppression des schwas soit légèrement plus marquée dans le cas des schwas interconsonantiques.

11.3.2 Diatopie de la syncope

Si, sur le plan de la dynamique diachronique, la situation des schwas internes interconsonantiques est analogue à celle des schwas internes hiatiques, il n'en va pas de même sur le plan diatopique. En effet, le large axe Nord-Ouest - Sud-Est que l'on pouvait déceler pour la réduction des hiatus internes est moins marqué pour la syncope, comme représenté sur la carte en fig. 11.7 page suivante, et ce pour plusieurs raisons. D'une part, l'Angleterre semble pratiquer très peu de syncopes (2% de formes syncopées), alors qu'elle était la région où la réduction des hiatus était la plus forte (46%). Par ailleurs, si le Nord et le Hainaut demeurent bien en dessous de la moyenne des régions (respectivement 5 et 0% de syncope, contre 18% en moyenne), ce n'est pas le cas de la Wallonie : alors que la Wallonie n'appliquait aucune réduction de hiatus dans nos données, elle dispose de taux de syncope qui sont parmi les plus élevés (55%). Si l'on prend en compte le nombre de textes de chaque région, le taux de la Wallonie est le seul taux supérieur à 35% à s'appuyer sur plus de 2 textes, ce qui en fait le seul taux très élevé à disposer d'une assise quantitative

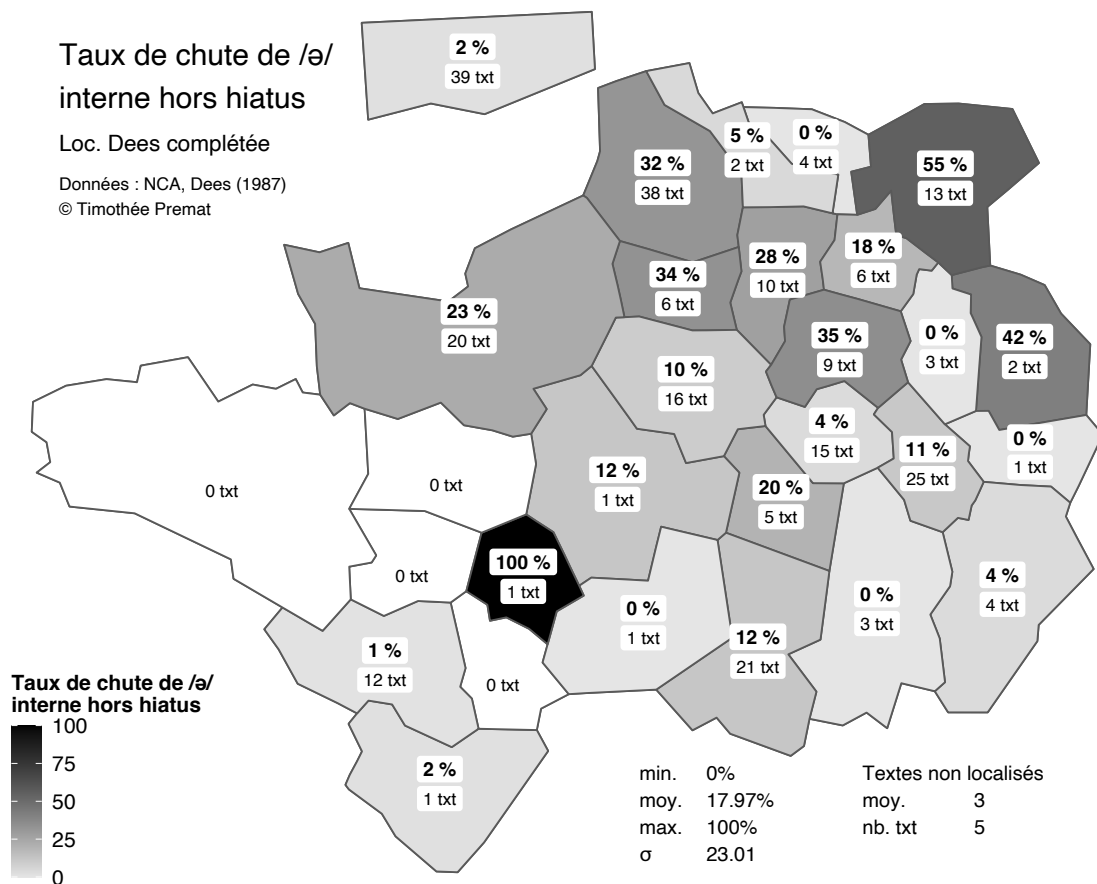


FIGURE 11.7 – Diatopie des taux de syncope, sur l'ensemble de la période considérée.

satisfaisante en termes de textes. Ainsi, il faut exclure de considérer trop sérieusement le taux de 100% de l'Indre-et-Loire.

Le panorama diatopique général qui se dégage de ces cartes synchroniques ne correspond donc pas à l'axe Nord-Ouest - Sud-Est que nous proposons de voir dans les cartes de la réduction des hiatus internes, parce que son pôle Nord-Ouest dominant, l'Angleterre, est tout simplement absent, et parce que le Nord-Est pratique la syncope, que ce soit en Wallonie (55%), en région Moselle et Meurthe-et-Moselle (42%), mais également dans les Ardennes (18%, contre 0% de réduction des hiatus). Par ailleurs, la Normandie pratique plus l'apocope que la moyenne des régions (23%), tandis que la Franche-Comté, qui représentait l'extrémité Sud-Est de l'axe de réduction des hiatus, pratique bien moins la syncope que la moyenne des régions (4%). Nous avons plutôt affaire, ici, à une distribution de la syncope assez forte au Nord du domaine, mais à l'exception de l'Angleterre, du Nord et du Hainaut, et qui devient de plus en plus faible à mesure que l'on progresse vers le Sud-Ouest.

11.3.2.1 En somme

Puisque les premières attestations graphiques de la syncope précèdent notre corpus (avec celles d'*Alexis*), il n'est pas étonnant que l'on trouve leurs premières attestations dans notre corpus dès 1125. La diffusion de la syncope dans le lexique est néanmoins progressive, le dernier lemme atteint l'étant à la fin du ^{xiii} s. Comme pour les hiatus internes, les formes syncopées n'évincent jamais totalement les formes non syncopées, mais le taux moyen de syncope (17%) est plus élevé que le taux moyen de réduction des hiatus (9%), et progresse de manière plutôt régulière à partir du début du ^{xiii} s.

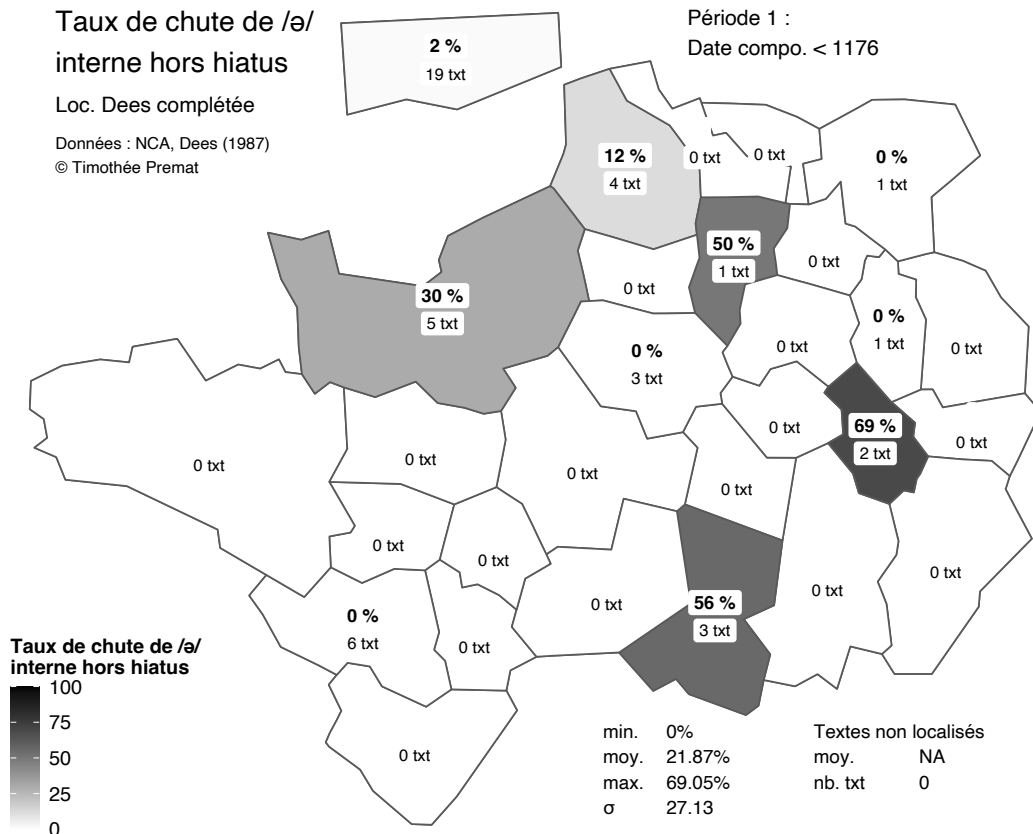
11.3.3 Intersection entre la diachronie et la diatopie

Pour étudier le développement diachronique de la diatopie, ou le développement diatopique de la diachronie, nous reproduisons ici la division en trois périodes appliquée pour les schwas internes en hiatus *supra*, pour obtenir les cartes en fig. 11.8 et 11.9, page suivante. Il convient d'emblée de noter que certaines caractéristiques des cartes synchroniques des syncopes sont transcrites telles quelles dans les cartes diachroniques. Ainsi, la Normandie pratique toujours des taux de syncope supérieurs ou égaux à la moyenne (de 17% en période 2 à 31% en période 3), tandis que l'Angleterre ne pratique que très peu de syncopes, avec un maximum à 3% en période 2. Cette situation est exactement inverse à celle que nous avons notée pour la réduction des hiatus internes, où l'Angleterre dominait, à toutes les périodes, en termes de réduction des hiatus internes, alors que la Normandie n'appliquait jamais cette réduction des hiatus internes.

11.3.3.1 1100-1175

Sur la première période (fig. 11.8.a), les régions qui répondent à nos huit lemmes sujets à syncope sont globalement les mêmes que celles qui répondaient à nos huit lemmes hiatiques, ce qui est en partie dû au fait qu'on trouve peu de textes sur cette période, et que ceux-ci sont plutôt conséquents. Sur cette période, on peut tenter de deviner, malgré l'absence de données dans plus de la moitié des régions, une distribution qui reprend l'axe Nord-Ouest - Sud-Est observé pour les schwas internes hiatiques (à l'exception, donc, de l'Angleterre, et en ajoutant la Normandie) : les taux de syncope supérieurs à 0% évitent le Nord-Est et le Sud-Ouest, mais également la région parisienne. De par l'absence générale de données, le tracé de cet axe n'est guère consolidé ; ne témoignent, par la négative, du tracé de l'axe, que les régions Vendée et Deux-Sèvres, région parisienne, Meuse et Wallonie. Il faut donc se garder de cette lecture qui reproduit l'axe Nord-Ouest - Sud-Ouest, dans la mesure où celui-ci est bien moins consolidé que pour les hiatus internes, et où l'inversion des taux de la Normandie et de l'Angleterre dilue la précision du tracé. Par ailleurs, le seul texte meusien et les trois textes de la Nièvre et de l'Allier antérieurs à 1176 inversent également leur valeurs : la Meuse était la seule région continentale à appliquer une forte réduction des hiatus mais n'applique aucune syncope avant 1176, tandis que la région Nièvre et Allier n'appliquait aucune réduction des hiatus internes

a



b

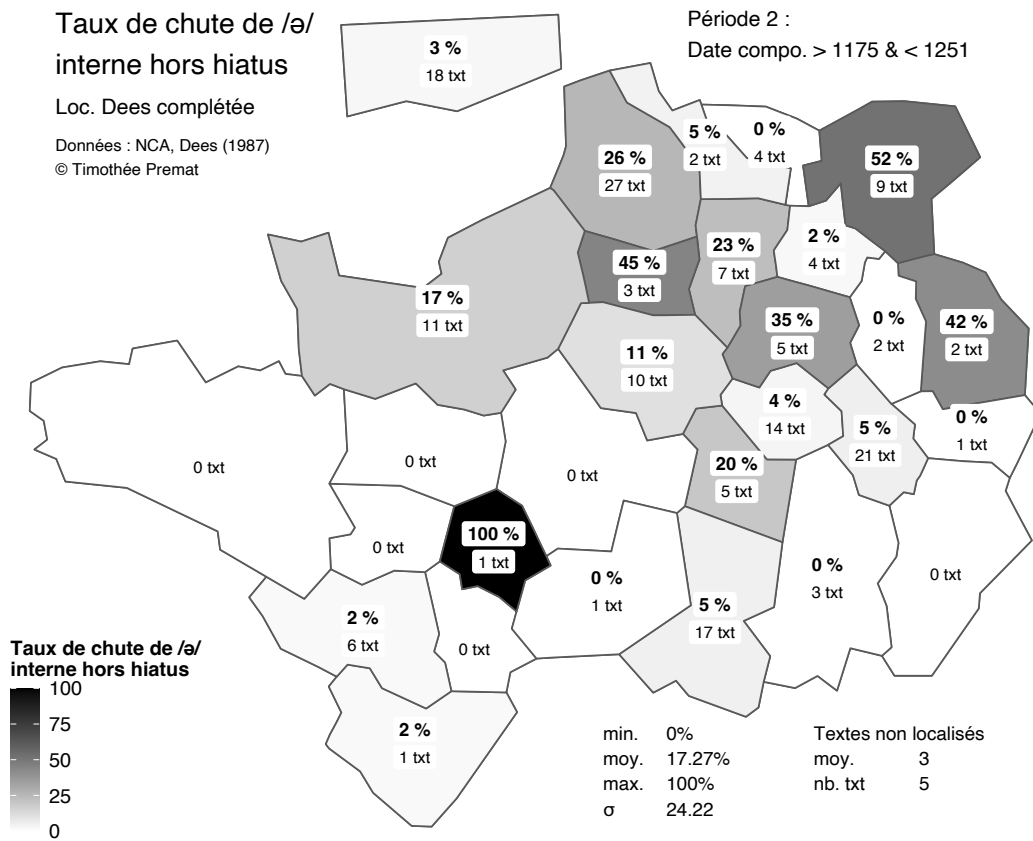


FIGURE 11.8 – Diatopie des taux de syncope dans le NCA, périodes 1 (1100-1176) et 2 (1176-1250).

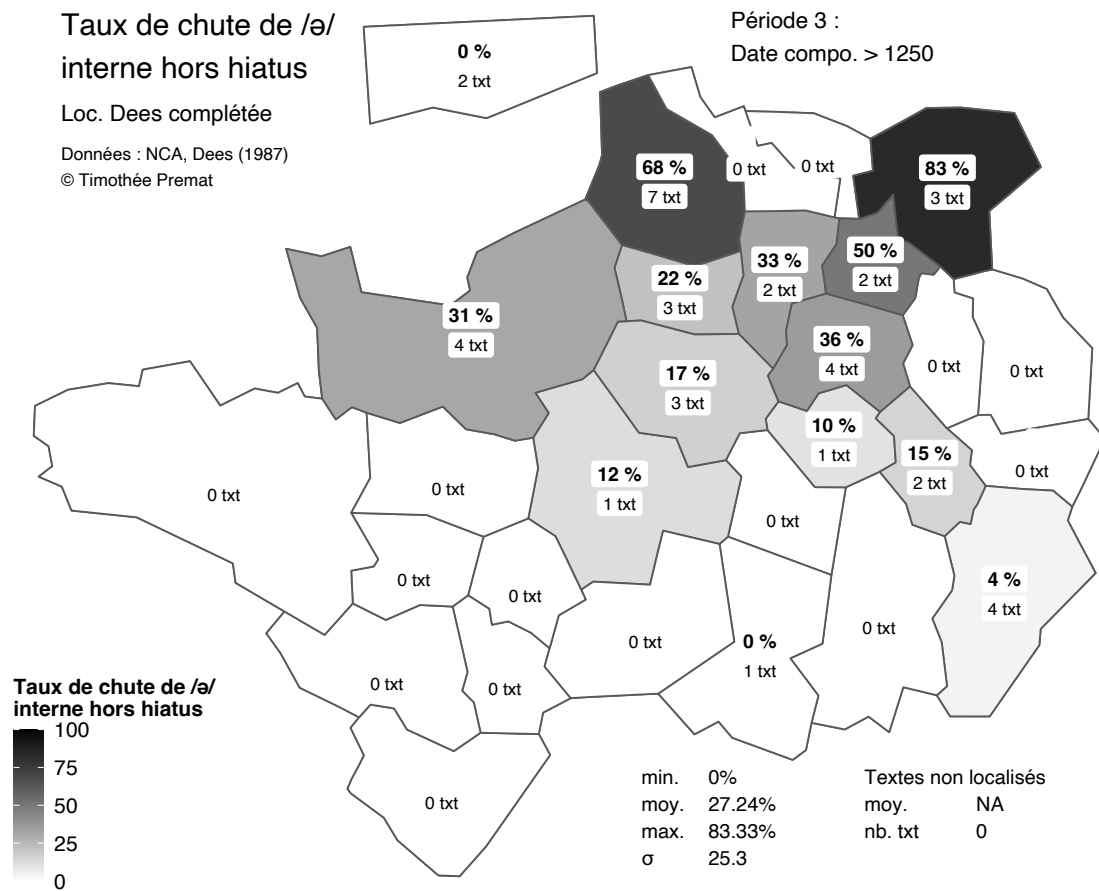


FIGURE 11.9 – Diatopie des taux de syncope dans le NCA, période 3 (1251-1322).

mais applique des taux conséquents de syncope (56%). Il en va de même pour la Haute-Marne, mais cette fois-ci le nombre de textes répondant à nos interrogations évolue : 4 textes antérieurs à 1176 répondaient à nos lemmes hiatiques mais ne pratiquaient aucune réduction de ces hiatus, tandis que seuls deux textes répondent à nos 8 lemmes syncopés, et pratiquent la syncope dans 69% des cas.

11.3.3.2 1176-1250

En période 2 (fig. 11.8.b), toute possibilité de lire un axe Nord-Ouest - Sud-Est analogue à celui de la réduction des hiatus internes disparaît : si l'on note certes une fine bande sans syncope ou à faible syncope constituée par le Nord, le Hainaut, les Ardennes, la Meuse et les Vosges, celle-ci est peu consolidée en termes de nombre de textes (jamais plus de 4 textes par région), tandis que la Wallonie et la région Moselle et Meurthe-et-Moselle adoptent fortement la syncope. L'évitement du Sud-Ouest du domaine est lui aussi mis à mal par l'Indre-et-Loire, mais le taux de 100% de syncope qui y est observé est peu fiable : non seulement il est géographiquement isolé, mais il ne tient qu'à un seul texte, qui

contient 6 formes syncopées et aucune forme non syncopée. La situation qui se dessine entre la fin du XII^e s. et la moitié du XIII^e s. correspond donc plutôt à une contamination générale des trois quarts du domaine, seuls le Sud-Ouest et l'Angleterre ne montrant que peu de syncopes.

11.3.3.3 1251-1322

Enfin, après 1250 (fig. 11.9), la contamination est générale sur le continent : toutes les régions qui répondent à nos huit lemmes disposent de syncopes, à l'exception de la région Nièvre et Allier, dont le seul texte postérieur à 1250, *mace*, dispose de 10 formes non syncopées.⁴² Ainsi, à l'exception de l'Angleterre de la région Nièvre et Allier, toutes les régions qui répondent à nos huit lemmes pratiquent la syncope. Néanmoins, cette pratique de la syncope varie de 4 à 83% d'occurrences syncopées, et la distribution semble ici significative : les régions aux taux inférieurs à la moyenne (27%) se situent toutes sous un tracé en arc de cercle défini par la Normandie, le Nord, l'Aisne et la Marne. De fait, toutes les régions disposant d'un taux supérieur à 20% se situent dans le tiers Nord du domaine et, hormis l'Angleterre, toutes les régions de ce tiers Nord disposent de taux supérieurs à 20%. Ainsi, à partir de la seconde moitié du XII^e s., le Nord du domaine montre une plus forte productivité de la syncope, tandis que celle-ci n'est que faiblement attestée dans le centre et le Sud du domaine, mais en partie en l'absence d'occurrences de nos lemmes dans ces régions.

mace : MACÉ DE LA CHARITÉ (1300*). *La Bible*. Dees : Nièvre, Allier; compo. Centre, ca 1300; ms. Berry, 1343; CR (MACÉ DE LA CHARITÉ 1977); DEAF : *BibleMacéV*.

42: Pour les lemmes « sairement », « porter- », « folement », « durement » et « esperon ».

11.3.3.4 En somme

Aucune de nos trois périodes ne reproduit donc précisément le tracé de l'axe Nord-Ouest - Sud-Est observé pour la réduction des hiatus internes. L'inversion du comportement de l'Angleterre et de la Normandie entre réduction des hiatus et syncope est notable, et est stable dans les trois périodes. Si les régions centrales de l'axe Nord-Ouest - Sud-Est observé pour la réduction des hiatus répondent généralement, dans les trois périodes, à la syncope, c'est principalement parce que la répartition de la syncope est tout simplement plus générale que celle de la réduction des hiatus internes. Enfin, l'exclusion de l'Angleterre, dans toutes les périodes, ne laisse pas d'interroger : c'est pourtant en Anglo-Normand que SHORT (2007, p. 95) situe les premières attestations de la syncope, avec *Alexis* (sec. 11.1.1.2). Les autres occurrences de la syncope que Short note dans différents textes ne semblent pas faire le poids, quantitativement, dans notre corpus : elles coexistent avec un nombre bien supérieur de formes non syncopées. Pourtant, l'anglo-normand est généralement décrit comme représentant une plus forte instabilité de schwas que le français continental (SHORT 2007, p. 39, 92-99; POPE 1966 [1934], p. 457). Il faut, sur ce point, noter que POPE (1966 [1934], p. 437) date l'instabilité générale de schwa en a.n., dans toutes les positions, de la moitié du XIII^e s., c'est-à-dire sur la période où les textes anglo-normands se font plus rares dans le NCA ; seuls deux textes répondent à nos huit lemmes à syncope potentielle en Angleterre après 1250. Par ailleurs, notre travail ne porte ici que sur huit lemmes, et si ce sondage permet d'avoir une première idée de la distribution du phénomène, et permet de premières attestations qui soient fiables en tant que telles, il faut se garder de juger

de sa valeur de représentativité sur l'ensemble des lemmes concernés par la syncope. Il n'en demeure pas moins qu'il est marquant que, sur nos huit lemmes à syncope potentielle, l'Angleterre est une région qui ne pratique que très peu ces syncopes.

11.4 Discussion

Dans cette section, nous regroupons les éléments exposés *supra* (sec. 11.4.1), et comparons la réduction des hiatus et la syncope (sec. 11.4.2), pour définir ce que les deux processus ont en commun (11.4.2.1) et ce qui les différencie (11.4.2.2). Ces différences nous amènent à conclure qu'il ne faut pas considérer ces deux phénomènes comme formant un tout homogène. Nous remarquons également que, que ce soit en diachronie ou en diatopie, ni la réduction des hiatus internes ni la syncope ne présentent une distribution qui ressemble à celles que nous avons notées pour les schwas finaux, ce qui nous amène à conclure que les processus de suppression des schwas internes semblent évoluer de manière plus ou moins distincte de celle des schwas finaux (sec. 11.4.2.3).

11.4.1 Synthèse

Les schwas internes sont sujets à deux processus de suppression durant la période littéraire de l'ancien français : la réduction des hiatus, qui concerne les schwas non finaux au contact d'une autre voyelle (sec. 11.2), et la syncope, qui concerne les schwas prétoniques précédés ou suivis d'une liquide ou d'une nasale (sec. 11.3).

11.4.1.1 Réduction des hiatus

Le premier phénomène de suppression des schwas est mal daté par la bibliographie; notre étude par sondage lemmatique a permis de montrer (sec. 11.2.1.2) que les premières occurrences de réduction des hiatus remontent à la première moitié du *xiii^e* s. Le phénomène n'est pas instantané, à l'instar de ce que supposerait une loi phonétique néogrammaticienne : il se diffuse progressivement dans le lexique. Pour les huit lemmes de notre sondage, la diffusion de la réduction des hiatus est achevée au milieu du *xiii^e* s.

Sur le plan diatopique (sec. 11.2.2), la réduction des hiatus internes est particulièrement forte en anglo-normand, et ce sur l'ensemble de la période. Elle semble se diffuser le long d'un axe Nord-Ouest - Sud-Est, mais celui-ci perd un peu de sa consistance après 1250, sans pour autant qu'on puisse avancer une dynamique émanant du centre du domaine, dans la mesure où l'Île-de-France conserve des taux de réduction de hiatus supérieurs à la moyenne (sec. 11.2.3).

11.4.1.2 Syncope

La syncope, elle, est attestée par la bibliographie dès le très ancien français. À ces attestations seulement graphiques, ou graphiques et métriques, peuvent s'ajouter des attestations non graphiques mais métriques, selon la position que l'on adopte vis-à-vis des vers anisosyllabiques. Bien qu'attestée dès avant notre corpus, la syncope suit le même processus de diffusion lexicale progressive : tous les lemmes ne sont pas atteints à la même date, et les premières syncopes des huit lemmes de notre corpus s'étendent de la première moitié du XII^e s. à la fin du XIII^e s. (sec. 11.3.1). Par ailleurs, certaines occurrences syncopées présentent un caractère exceptionnel : ainsi, *portra*, tout en étant bien formé sur le plan phonotactique, ne dispose que d'une seule occurrence dans notre corpus, la base *porter*-résistant très bien à la syncope. L'influence paradigmatique, ici, ne peut être avancée qu'à titre anecdotique : les formes *donra*, par exemple, sont bien mieux attestées et sont un exemple classique repris par plusieurs grammaires. La diffusion lexicale est donc inégale à plusieurs titres : non seulement en termes de première attestation par lemme, mais aussi en termes de maintien de la productivité.

Sur le plan diatopique (sec. 11.3.2), la diffusion de la syncope est bien plus homogène que celle de la réduction des hiatus, puisqu'elle concerne l'ensemble du domaine à l'exception de l'Angleterre et du quart Sud-Ouest, où elle demeure marginale ou mal attestée par manque de textes. La diachronie de cette diatopie ne présente guère d'effet marqué : la syncope apparaît partout où des textes attestent des lemmes de notre sondage, sauf en Angleterre et dans le Sud-Ouest (sec. 11.3.3). Il faut néanmoins remarquer la bande constituée par le Nord, le Hainaut, les Ardennes, la Meuse et les Vosges en période 2 (entre 1175 et 1250), où la syncope est très peu attestée. Dans la mesure où cette bande n'inclut pas la Wallonie et la région Moselle et Meurthe-et-Moselle, cette distribution ne saurait défendre l'existence de l'axe Nord-Ouest - Sud-Est constaté pour la réduction des hiatus.

11.4.2 Points communs et différences

Cette section discute les points communs et les différences, d'une part entre la réduction des hiatus internes et la syncope (sec. 11.4.2.1 et 11.4.2.2), et d'autre part entre ces deux phénomènes et les défauts de voyelles finales atones (sec. 11.4.2.3). Elle aboutit à la conclusion qu'il ne paraît pas adapté d'essayer de fondre ces trois phénomènes dans une analyse commune qui procéderait simplement d'une élimination de ces voyelles.

11.4.2.1 La coexistence des graphies en <e> et sans <e> est commune

Pour la réduction des hiatus internes comme pour la syncope, et conformément à ce que note la bibliographie, les formes dont le schwa interne n'est pas supprimé ne disparaissent pas ; loin de là, elles demeurent majoritaires sur la plupart des régions et sur l'ensemble de la période (sec. 11.2.1.3 et 11.3.1.3). Dans les deux cas, on trouve bien plus de textes n'ayant aucune forme à schwa supprimé que de textes n'ayant que des formes à schwa supprimé, selon un rapport de 1 pour 25 pour la réduction

des hiatus, et de 1 pour 15 pour la syncope. Le taux moyen de réduction des hiatus est plus faible que celui de syncope, à 9% de hiatus réduits contre 17% de syncope; mais notre approche par sondage lemmatique ne garantit pas la représentativité de ces chiffres.

Par ailleurs, ces données ne sont que graphiques, or il est possible de considérer que la conservation graphique d'un hiatus n'est qu'une marque de longueur vocalique (sec. 11.1.1.1, sous 11.8) et que, si l'on voulait accorder à la métrique le statut de témoin de phénomènes phonologiques, on pourrait avancer qu'il existe des occurrences de syncope non transcrites (*trovereiz* bisyllabique à côté de *durai* pour *duneraiz* dans *Alexis* en 11.11, sec. 11.1.1.2). Mais la métrique, si elle peut être considérée comme capable de révéler des chutes de schwa non transcrites, est aussi capable de conserver des schwas qui seraient tombés dans la langue non versifiée, la paraphonologie métrique procédant souvent, au moins en partie, d'un conservatisme linguistique. S'il n'y a donc, ici, pas de conclusion définitive possible, il convient néanmoins de conserver une certaine précaution vis-à-vis des occurrences dont la graphie demeure hiatique ou non syncopée, dans la mesure où celle-ci peut masquer une chute du schwa.

11.4.2.2 La motivation phonologique ne peut être commune

Si la réduction des hiatus internes et la syncope semblent suivre une dynamique globale commune, il faut néanmoins noter que cette dynamique ne suit pas la même distribution diatopique : la ressemblance observée sur les graphiques de séries temporelles en fig. 11.1 et 11.6 cache un développement diatopique différencié de la réduction du hiatus et de la syncope. De ce fait, il ne convient pas de considérer que les deux phénomènes seraient le produit d'un même processus phonologique, dans la mesure où une motivation unique prédirait un développement diatopique semblable dans les deux cas.

Par ailleurs, une motivation unique semble difficile à accepter du point de vue phonotactique : la seule chose que la réduction des hiatus et la syncope ont en commun, c'est de cibler un schwa; et encore, dans les hiatus, la voyelle en question peut être notée par un autre graphème que <e>. Du point de vue phonotactique, les deux processus sont opposés : la réduction des hiatus supprime une structure marquée (CVVC) au profit d'une structure moins marquée (CVC ou CV:C), tandis que la syncope remplace une structure non marquée ou peu marquée (CVC) par une structure plus marquée (CC), créant des groupes coda-attaque, des attaques branchantes et des groupes coda-attaque branchante. Non seulement il est difficile d'imaginer une motivation phonologique commune à ces deux comportements contradictoires, mais, s'il existait une motivation phonotactique commune aux deux phénomènes, on s'attendrait également à ce qu'elle favorise la réduction des hiatus plus que la syncope, et non l'inverse. De fait, la phonotactique ne peut, seule, prédire la réduction des hiatus et la syncope, et ne peut non plus prédire une distribution diatopique différenciée. La réduction des hiatus internes et la syncope ne peuvent donc pas être réduits à un processus phonotactique qui serait homogène.

11.4.2.3 Comparaison avec les schwas finaux

Les données utilisées pour traiter de la réduction des hiatus internes et de la syncope, si elles ne sont pas de la même ampleur que celles utilisées pour les voyelles finales atones, sont également fort différentes. Aucune des parties du discours étudiées dans ce travail ne témoigne d'une évolution diachronique proche de celle de la réduction des hiatus internes ou de la syncope. Les taux de défauts des schwas finaux sont nuls ou marginaux pour les noms et les adjectifs, et sont bien plus élevés pour les adverbes et les monosyllabes fonctionnels. La distribution diatopique de la réduction des hiatus et de la syncope ne coïncide pas non plus avec les distributions diatopiques observées pour les voyelles finales atones, d'une part parce que réduction des hiatus internes et syncope ne suivent pas la même distribution diatopique, et d'autre part parce que l'évolution diatopique est divergente pour le schwa final en fonction des parties du discours voire des lemmes. Dans certains cas, les distributions diatopique s'opposent clairement; ainsi, alors que la réduction des hiatus est principalement attestée sur un axe Nord-Ouest - Sud-Est, c'est précisément sur cet axe que les conjonctions conservent le mieux leur <e> final (devant initiale vocalique pour le *que* des locutions en *base* + *-que* et de l'élément *que* et dans les deux contextes pour « come », cf. chap. 10).

Par ailleurs, le fait que la réduction des hiatus et la syncope ne se diffusent que progressivement dans le corps lexical marque une autre différence avec les données observées pour le schwa final. En effet, pour la majeure partie des lemmes sujets au défaut de leur voyelle finale, celle-ci ne montre pas de diffusion progressive dans le lexique. Si l'on voulait placer les réductions de hiatus internes et les syncopes dans le même paradigme que les cas de défauts de voyelles finales atones, il faudrait expliquer cette différence, qui est essentielle : pour les schwas finaux, la phonologie se préoccupe peu du lemme, alors que pour les hiatus internes et les syncopes, elle progresse différemment selon le lemme en question. Ces différences dans les taux d'absence de graphème, dans la distribution diachronique et diatopique de ces taux, et dans la façon dont les lemmes sont concernés, nous invitent à considérer que la réduction des hiatus internes et la syncope ne peuvent être considérés comme procédant exactement comme les défauts de voyelles finales atone.

11.4.3 En somme

Ainsi, si la réduction des hiatus internes et la syncope ont en commun une certaine dynamique diachronique, elles se distinguent par leur dynamique diatopique et par la marque de la configuration phonologique qu'elles génèrent. De ce fait, il n'est pas possible de considérer que les deux suppressions de schwa procèdent d'une simple contrainte phonotactique, dans la mesure où si la réduction du hiatus va dans le sens de la phonotactique, la syncope crée au contraire des configurations plus complexes. Cette différence diatopique et phonotactique nous invite à considérer que, malgré ce que ces deux processus ont en commun, ils ne sauraient être considérés de manière unifiée.

Par ailleurs, ces deux phénomènes de disparition du schwa interne se distinguent assez clairement des données dont nous disposons pour le

schwa final : ni la diatopie, ni la diachronie, ni la diffusion lexicale ne coïncident. Il convient donc de postuler que non seulement la réduction des hiatus internes et la syncope ne représentent pas un seul et même processus, mais aussi d'avancer que leur développement est en apparence orthogonal à celui des chutes de schwa final, et donc, à celui de l'élision.

11.5 Conclusion

Dans ce chapitre, nous nous sommes intéressés à deux autres processus de suppression de schwa en ancien français : la réduction des hiatus internes et la syncope. Après avoir décrit ces deux phénomènes à partir de la bibliographie (sec. 11.1.1.1 et 11.1.1.2), nous avons entrepris de documenter précisément chaque phénomène à partir d'un sondage portant sur huit lemmes. Nous avons ensuite dédié une section à chaque phénomène (sec. 11.2 et 11.3), en détaillant dans chaque section la distribution diachronique (sec. 11.2.1 et 11.3.1) et diatopique (sec. 11.2.2 et 11.3.2) ainsi que l'intersection de ces deux distributions (sec. 11.2.3 et 11.3.3). Une fois cette description effectuée, nous avons discuté ces résultats (sec. 11.4), pour établir que si la réduction des hiatus internes et la syncope partagent une dynamique diachronique et une coexistence avec les graphies où <e> est conservé communes (sec. 11.4.2.1), elles ne pouvaient néanmoins émaner d'une même motivation phonotactique, dans la mesure où leur relation à la phonotactique est opposée et où leurs distributions diatopiques ne concordent pas (11.4.2.2). Nous avons également avancé que ces deux processus produisent des distributions différentes de celles observées pour les voyelles finales atones, que ce soit sur le plan quantitatif, sur le plan diatopique ou sur le plan diachronique, ce qui nous a amené à considérer qu'il s'agissait d'évolutions au moins partiellement différenciées (sec. 11.4.2.3).

III

ANALYSE

LA dernière partie de ce travail est organisée selon quatre chapitres. Le premier (chap. 12) reprend l'ensemble des données graphiques développées dans la partie précédente et en tire des généralisations d'abord descriptives puis analytiques. Le deuxième chapitre (chap. 13) compare ces généralisations avec celles que l'on peut tirer des sources métriques et musicales, ainsi qu'avec le discours, postérieur, des grammairiens du xvi^e s. Le troisième chapitre (chap. 14) réunit les données graphiques et les données métriques, musicales et métalinguistiques, et propose une analyse phonologique du phénomène de l'élosion. Enfin, le dernier chapitre (15) conclut ce travail en résumant la teneur, les apports et les limites.

Ce chapitre vise à réunir les données décrites dans la partie précédente, à les comparer et à tirer de cette comparaison les généralisations pertinentes pour la description et pour l'analyse phonologique des phénomènes de chute des voyelles finales atones. Il commence par une synthèse des taux de défauts (sec. 12.1); à partir des taux de défaut par catégorie, nous dérivons trois typologies, selon que la variable utilisée pour le classement est le taux absolu (sec. 12.1.1), le rapport entre taux de défaut devant initiale vocalique et celui devant initiale consonantique (sec. 12.1.2) ou la puissance de l'effet du contexte tel que mesuré par le test de Student (sec. 12.1.3). Dans la deuxième section, c'est par le croisement entre ces différentes entrées typologiques que nous parvenons à mettre au point la typologie qui rend le mieux compte de nos données (sec. 12.2, ce qui nous permet de définir quels cas d'absence de <e> tiennent au processus de l'élision, et lesquels procèdent d'une autre logique. Dans la troisième section, nous parvenons à décrire les caractéristiques majeures de nos données sous la forme de deux (sec. 12.3) puis d'une condition principale à l'élision (sec. 12.3.1), qui s'accompagne d'une condition qui rend compte d'un sous-ensemble marginal d'élisions motivées par des collocations spécifiques (sec. 12.3.2).

12.1 Synthèse des taux de défaut

L'ensemble des données concernant les défauts de voyelles finales atones, ainsi que pour les schwas internes, est reproduit en table 12.1. Sur cette table, les catégories sont données dans l'ordre dans lequel elles ont été exposées dans la partie précédente. La colonne *Taux.Ø/Nb.* correspond au nombre de formes défectueuses rapporté au nombre total de formes de cette catégorie, tandis que la colonne *Tx_Ø* correspond au taux de formes défectueuses uniquement pour les lemmes qui sont concernés par une alternance <e,V>/Ø; la première de ces deux colonnes n'est donc pas renseignée lorsque la catégorie correspond à un seul lemme — elle est indiquée pour les conjonctions en *base + -que* puisque cette catégorie recouvre plusieurs lemmes. Les valeurs *t* correspondent à celles du test de Student mesurant l'effet du contexte effectué par *Comparalem*; la valeur *p* n'est indiquée que lorsqu'elle est supérieure à 0,05. Enfin, la dernière colonne indique le taux de défaut devant initiale vocalique.

La lecture de la table 12.1 permet de classer les catégories selon plusieurs entrées typologiques, en fonction des variables sélectionnées pour le classement. Dans ce qui suit, nous ébauchons une première typologie en fonction des taux absolus de défaut (sec. 12.1.1), puis affinons le classement à l'aide des taux de défaut dans chaque contexte (sec. 12.1.2) puis de l'effet du test de Student sur l'influence du contexte (sec. 12.1.3). La combinaison de ces trois approches nous permet, dans la section suivante (sec. 12.2), de parvenir à la typologie finale des cas d'élision en

TABLE 12.1 – Synthèse des défauts de <e> ou <V> par catégorie. Les valeurs *t* et *p* sont celles du test de Student mesurant l'effet du contexte sur les taux d'absence de graphème vocalique.

Chap./sec.	Catégorie	Taux.Ø/Nb.	Tx_Ø	<i>t</i>	(<i>p</i>) ^a	Tx_Ø_#V
4	NOM	< 0,01%	0,68%	0,49	(0,6)	0,01%
5	ADJ	0,57%	2,43%	4,84		6%
6	PRÉP ^b	4,48%	12,88%	85,92		68%
7	ADV ^c	5,78%	75,08%	-2,5		87%
9.2.3	PROPersMASC CR3sg	<i>le</i> ^d	55,07%	58		99%
9.2.4	PROPers CR2sg	<i>te</i>	35,24%	94,57		96%
9.2.5	PROPers CR1sg	<i>me</i>	44,65%	155,3		97%
9.2.6	PROPers CS1sg	<i>jo</i>	7,73%	21,57		48%
9.2.7	PROPers CS2sg	<i>tu</i>	0,28%	3,1		1%
9.2.8	PROPers FÉM CS3sg	<i>ele</i>	10,92%	-4,15		6%
8.2.2	ART DÉF MASC CRsg	<i>le</i>	6,78%	350,34		99%
8.2.2	ART DÉF FÉM CS-Rsg	<i>la</i> ^e	15,95%	531,52		97%
8.2.3	ART DÉF MASC CSsg	<i>li</i>	4,43%	17		33%
8.2.3	ART INDÉF FÉM CS-Rsg	<i>une</i>	2,21%	5,65		9%
8.2.4	ART DÉF MASC CSpl	<i>li</i>	< 0,01%	1,68	(0,09)	0,01%
10.2.2	CONJ <i>base</i> + <i>-que</i>	0,49%	20,16%	26,48		69%
10.2.3	CONJ élément <i>que</i>		9,63%	32		41%
10.2.4	CONJ « come » ^f		60,3%	-1,88	(0,06)	62%
11.2	Hiatus interne		9%			
11.3	Syncope		17%			

a. Uniquement si $p > 0,05$ (mesure *t* de Student non significative); $p < 0,05$ pour les Adv et $< 0,005$ pour les autres catégories.

b. Sauf « (de-)sor ».

c. Sauf « ensemble ».

d. Les formes ici présentées ne sont que les formes canoniques, elles n'excluent pas les variations graphiques relevées dans les chapitres de la section II.

e. Sauf formes picardes *le*, *li*.

f. Tous contextes syntactiques.

fonction de leurs valeurs absolues, contextuelles, et d'autres variables (temps, espace, contexte morphosyntaxique).

12.1.1 Typologie selon les taux absolus de défaut

Une première typologie peut être esquissée, selon les taux absolus de défaut. Celle-ci est représentée en (12.1), où les seuils entre les différentes catégories sont certes arbitraires mais permettent de visualiser les taux de défaut des différentes catégories. Nous proposons, ici, de séparer les catégories dont plus de la moitié des formes sont défectueuses (12.1 a), celles dont une partie conséquente mais minoritaire des formes sont défectueuses (12.1 b), celles dont une partie moins conséquente mais toujours importante des formes sont défectueuses (12.1 c), celles dont les défauts sont marginaux mais néanmoins bien attestés relativement aux lemmes en question (12.1 d), et celles dont les défauts sont trop marginaux pour être considérés sérieusement (12.1 e). Pour les pronoms et les articles, les formes canoniques sont données à droite comme simple aide à la lecture.

Cette typologie suivant les taux absolus de défauts sur les lemmes considérés n'est pas inintéressante, dans la mesure où elle regroupe

(12.1) (a)	Ø > 50%	ADV	75%	
		CONJ « come »	60%	
		PROPersMASC CR3sg	55%	<i>le</i>
(b)	Ø > 35%	PROPers CR1sg	45%	<i>me</i>
		PROPers CR2sg	35%	<i>te</i>
(c)	Ø > 10%	CONJ <i>base + -que</i>	20%	
		ARTDéfFÉM CS-Rsg	16%	<i>la</i>
		PRÉP	13%	
		PROPersFÉM CS3sg	11%	<i>ele</i>
(d)	Ø > 1%	CONJ élément <i>que</i>	10%	
		PROPers C1sg	8%	<i>je, jo</i>
		ARTDéfFÉM CS-Rsg	7%	<i>la</i>
		ARTDéfMASC CSsg	4%	<i>li</i>
		ADJ	2%	
		ARTIndéfFÉM CS-Rsg	2%	<i>une</i>
(e)	Ø < 1%	NOM	0,68%	
		PROPers CS2sg	0,28%	<i>tu</i>
		ARTDéfMASC CSpl	< 0,01%	<i>li</i>

dans la catégorie supérieure de défauts (de 50 à 75%) les adverbes, les conjonctions « come » et les pronoms personnels de troisième personne du singulier au cas régime direct. Ces trois catégories sont pourtant hétérogènes : les adverbes ne connaissent qu'un faible effet du contexte, en faveur du contexte préconsonantique ($t = -2,5$; sec. 7.2.4) ; le lemme « come » ne correspond pas à strictement parler à ce que nous qualifions de *défaut de <e>* (sec. 10.1.3), et le pronom personnel *le* est le seul à connaître massivement l'enclise avec un terme support précédent (sec. 9.2.3.1). Nous avons déjà évoqué le rapprochement entre ces deux premières catégories lors du développement concernant « come » (sec. 10.2.4 et 10.3.2), où nous notions que la forme attendue de ce mot est *a priori* du type *com*, tandis que le type concurrent, *come*, est originellement restreint à l'introduction de propositions averbales. Il est possible, ici, de proposer l'hypothèse inverse : ce n'est pas « come » qui se comporte comme les adverbes, mais les adverbes qui se comportent comme « come ». Dans cette hypothèse, le clair détachement des adverbes vis-à-vis de nos autres catégories (taux de défaut maximal, faible influence, et influence inverse, du contexte) reviendrait à disqualifier la présence stable d'un schwa final dans la représentation sous-jacente des adverbes. Nous ne sommes pas en mesure d'apporter de réponse définitive à cette question, et laissons donc cette hypothèse en suspens, pour nous concentrer sur les cas d'élision. Quant aux pronoms personnels masculins au cas régime de la troisième personne du singulier, leur présence dans ce groupe s'explique donc par l'enclise qui provoque la réduction de *le* à *l* devant initiale vocalique à hauteur de 21% (moyenne des taux des textes). Cette enclise, combinée à une élision quasiment systématique devant initiale vocalique (99%), donne à ce pronom personnel un taux non contextuel de défaut bien supérieur à celui de tous les autres pronoms.

À l'opposée du spectre, les noms, pronoms personnels au cas sujet de la deuxième personne du singulier *tu* et les articles définis masculins au cas sujet pluriel *li* ont des taux de défauts marginaux (inférieurs à 1%), et même rarissimes pour l'article *li* pluriel (< 0,01%). On peut donc

considérer, face à ces données, que ces catégories ne sont concernées ni par l'élosion (devant #V) ni par l'apocope (devant #C), et que leur voyelle finale est stable. Dans ce cas, ces catégories témoignent d'un *bruit dans les données* plutôt que d'un *bruit de fond* qui serait doté d'une pertinence linguistique ; ce bruit dans les données peut être dû soit au corpus, qui demeure imparfait sur l'annotation et/ou la transcription de certaines formes, soit à la graphie médiévale elle-même, dont personne n'attend une régularité absolue.

Entre ces deux extrémités du spectre, on trouve des catégories dont les défauts sont plus ou moins rares mais auxquels on peut, selon les cas, donner une certaine signification linguistique. Ceux-ci peuvent demeurer à des taux faibles (12.1 d) — de 2% (pour les adjectifs et *une*) à 10% (élément *que*), plus forts mais encore largement minoritaire (12.1 c) — de 11% (*ele*) à 20% (CONJ en *base + -que*), ou minoritaires mais se rapprochant des 50% (35% pour *te*, 45% pour *me*).

De fait, ces différents taux de défaut non contextuels représentent un continuum, et les seuils que nous avons insérés dans ce continuum ne servent qu'à simplifier l'exposition et à préparer la typologie plus fine présentée *infra*. Le continuum de ces taux de défaut est représenté en fig. 12.1, où la dimension arbitraire des seuils que nous avons placés en (12.1) est évidente.

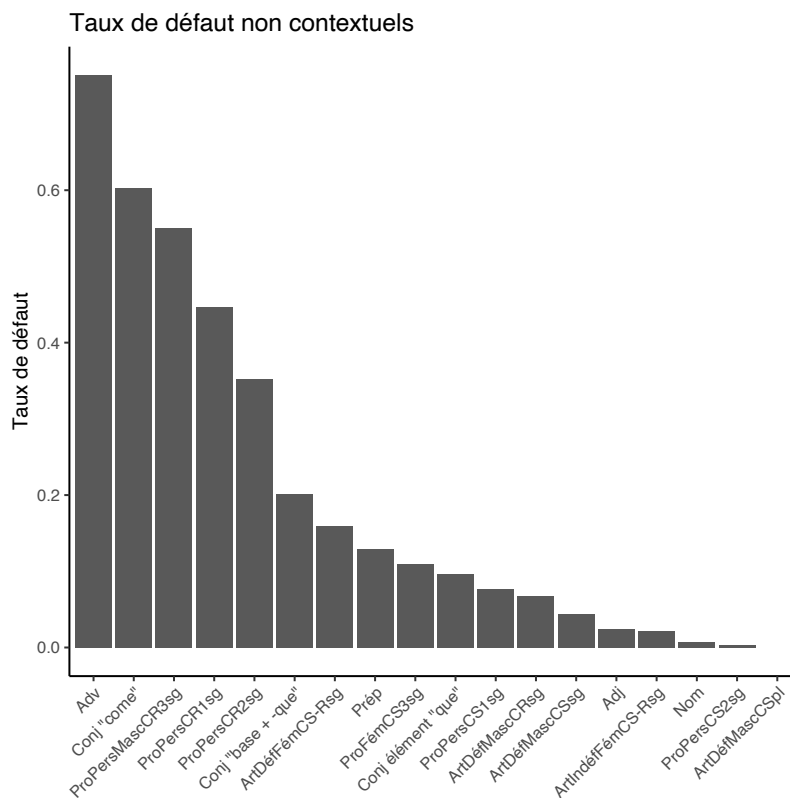


FIGURE 12.1 – Taux de défaut de voyelles finale atone selon les différentes catégories.

12.1.2 Typologie selon les taux contextuels de défaut

Comme indiqué précédemment, les trois catégories dont les taux de défaut sont les plus élevées (Adv, « come » et ProPers *le*) sont des catégories qui ne connaissent qu'une faible variation en fonction du contexte phonologique subséquent. De fait, la typologie proposée *supra* ignore tout de la distinction entre élision et apocope. Cette distinction peut être approchée par la comparaison des taux de défaut devant initiale vocalique avec les taux de défaut devant initiale consonnantique, ou, d'une manière moins simple à lire mais plus rigoureuse, à l'aide de la valeur-*t* du test de Student mesurant l'effet du contexte. La première présentation, mettant simplement en regard taux de défaut devant initiale vocalique et devant initiale consonnantique, est illustrée en fig. 12.2.

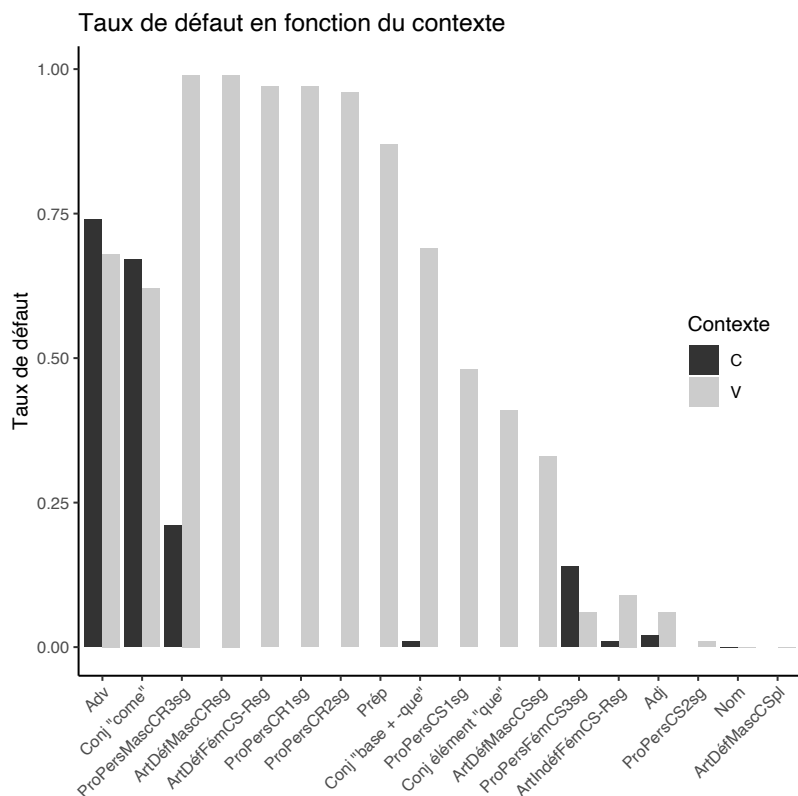


FIGURE 12.2 – Taux de défaut de voyelles finale atone selon les différentes catégories et le contexte subséquent.

En fig. 12.2, on peut noter que les adverbes et « come », non seulement partagent des valeurs relativement proches, mais ont également en commun d'avoir plus de défauts devant initiale consonnantique que vocalique, alors que c'est l'inverse pour toutes les autres catégories sauf le pronom *ele*. Une autre catégorie dispose également de taux de défaut plus élevés devant initiale consonnantique que devant initiale vocalique, les pronoms personnels sujets féminins singuliers *ele*; cette ressemblance dans la relation entre les valeurs des taux contextuels avait été notée en sec. 9.2.8.1. Quant aux pronoms personnels masculins au cas régime de troisième personne du singulier *le*, s'ils connaissent un taux de défaut devant initiale consonnantique important (21%), celui-ci est sans commune mesure avec celui des adverbes (74%) et de « come » (67%), et s'explique par le processus spécifique de l'enclise comme expliqué *supra*.

Ce qui apparaît le plus clairement en fig. 12.2, c'est l'existence d'un groupe de catégories qui ne connaissent le défaut que devant initiale vocalique, ou qui ne le connaissent que très rarement devant initiale consonnantique. Comme noté dans les chapitres de la partie précédente, ce comportement concerne, par intensité décroissante de l'élision, les articles définis masculins CRsg *le* et féminin CS-Rsg *la* (sec. 8.2.2), les pronoms personnels régimes singuliers de première *me* (sec. 9.2.5) et deuxième personnes *te* (sec. 9.2.4), les prépositions (chap. 6), les conjonctions en *base + -que* (sec. 10.2.2), le pronom personnel sujet de première personne du singulier *je, jo* (sec. 9.2.6), l'élément *que* (sec. 10.2.3) et l'article défini masculin CSsg *li* (sec. 8.2.3).

Enfin, le pronom personnel sujet féminin singulier *ele* (sec. 9.2.8), l'article indéfini féminin singulier *une* (sec. 8.2.3) et les adjectifs féminins singuliers (chap. 5) connaissent des défauts dans les deux contextes. Ces défauts sont plus nombreux devant initiale vocalique que devant initiale consonnantique pour *une* et pour les adjectifs féminins singuliers, tandis que le pronom personnel sujet de deuxième personne du singulier *tu* (sec. 9.2.7) ne connaît de défaut que devant initiale vocalique, mais marginalement (1%). Si nous avons proposé de considérer que le rapport entre les taux contextuels rapprochait *ele* des adverbes, il faut également avancer que ce rapport rapproche *une* des adjectifs féminins singuliers.

12.1.3 Typologie selon l'effet du contexte

Une autre manière d'approcher le rôle du contexte est d'utiliser la valeur-*t* du test de Student effectué pour mesurer l'effet du contexte sur les défauts de voyelle finale atone.¹ Si cette méthode propose des résultats moins immédiatement lisibles, elle a l'avantage de mesurer l'importance de la différence entre les deux contextes à l'aune de l'étendue de la variation de chaque contexte, ce qui génère également une valeur-*p* indiquant la probabilité que cette distribution soit due au hasard. Les valeurs-*t* sont représentées en fig. 12.3. La valeur-*t* est négative lorsque c'est le contexte préconsonnantique qui dispose du taux de défaut le plus élevé, et positive lorsque c'est l'inverse. Plus la valeur s'éloigne de zéro plus la différence, étant donnée la distribution, est conséquente, et les trois catégories pour lesquelles cette valeur-*t* pourrait être due au hasard ($p > 0,05$; l'hypothèse nulle ne peut être rejetée) sont indiquées en gris et non en noir. Du fait de l'étendue très forte des valeurs-*t*, l'échelle des ordonnées en fig. 12.3 est logarithmique.

La représentation en fig. 12.3 est donc plus rigoureuse que celle en fig. 12.2 *supra*. Les trois catégories dont le contexte préconsonnantique a un taux de défaut plus élevé devant initiale consonnantique que devant initiale vocalique disposent naturellement d'une valeur-*t* négative, et toujours proche de zéro (CONJ « come » : $t = -1,88$; ADV : $t = -2,5$; PROPERSONSFÉMCS3sg *ele* : $t = -4,15$). La valeur-*p* est trop élevée pour la conjonction « come » ($p = 0,06$), ce qui signifie que l'hypothèse nulle (le contexte n'a pas d'effet sur les taux de défaut) ne peut être rejetée. Deux autres catégories sont concernées par cette valeur-*p* trop élevée, les noms ($t = 0,49$; $p = 0,6$) et l'article défini masculin pluriel *li* ($t = 1,68$; $p = 0,09$). Ces deux dernières catégories ne comportant d'occurrences défectueuses que de manière exceptionnelles et probablement à cause d'un bruit dans les données, il n'y a rien d'étonnant à ce que l'hypothèse nulle ne puisse être rejetée.

1: Toutes les valeurs indiquées ici et *infra* sont celles que Comparalem a calculées et qui sont imprimées sur les sorties graphiques correspondant aux catégories, présentées dans les différents chapitres de la partie précédente.

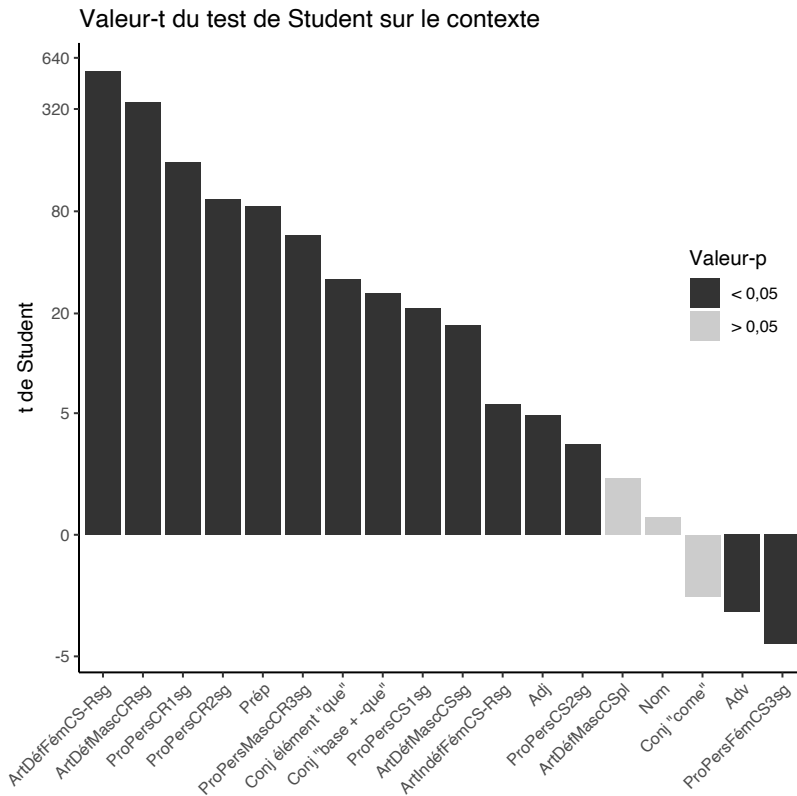


FIGURE 12.3 – Valeur- t du test de Student portant sur l'effet du contexte pour chaque catégorie.

12.1.4 En somme

On peut classer les catégories selon leur taux absolu de défaut, selon la différence entre leurs taux contextuels de défaut ou selon la force de l'effet du contexte. Ces différents classements amènent à faire des rapprochements entre les catégories, qui ne sont pas constants selon la variable adoptée. Dans la section suivante, nous fusionnons ces différentes approches pour définir ce qui tient de l'élosion et quel degré de force celle-ci a pour les différents groupes de catégories.

12.2 Ce qui tient de l'élosion, et ce qui n'en tient pas

Dans cette section, nous tentons d'établir les seuils dans nos mesures qui discriminent les élosions des autres processus à même de supprimer des voyelles finales atones, de manière à réduire la marge d'arbitraire d'une telle définition numérique, ou du moins à la justifier autant que possible. Nous parvenons à une définition explicite, sur le plan quantitatif, des différentes configurations de l'élosion, en étudiant l'interface entre les valeurs- t et les taux contextuels (sec. 12.2.1), avant de discuter ces catégories non plus sous l'aspect statistique mais du point de vue linguistique (sec. 12.2.2).

12.2.1 Interface entre valeurs-*t* et taux contextuels

Il va de soi que les catégories où les défauts sont plus nombreux devant initiale consonantique que devant initiale vocalique ne peuvent être considérés comme tenant de l'élision : si une suppression des voyelles finales est à l'œuvre autant voire plus devant initiale consonantique que devant initiale vocalique, il n'est pas raisonnable d'attribuer les suppressions devant initiale vocalique au processus contextuel de l'élision. C'est donc le cas du pronom *ele*, des adverbes et de la conjonction « come ». De même, l'exclusion des catégories pour lesquelles l'effet du contexte n'est pas assez fort pour que l'hypothèse nulle soit rejetée s'impose, ce qui place les noms et l'article défini masculin pluriel *li* en dehors du champ de l'élision. Néanmoins, un tri reste à opérer au sein des catégories pour lesquelles le contexte prévoicalique est significativement associé à plus de défauts que ne l'est le contexte préconsonantique (lorsque $t > 0$ et $p < 0,05$), étant donnée la variété des valeurs-*t* et des différences entre les taux contextuels.

Du fait de son échelle logarithmique, la fig. 12.3 donne l'impression d'une droite reliant les différentes valeurs, ce qui rend très difficile de détecter d'éventuels seuils dans ces données. Sans l'échelle logarithmique, les données prennent naturellement la forme d'une fonction exponentielle, tel qu'illustré en fig. 12.4, ce qui ne permet pas non plus un découpage plus fin. Néanmoins, le croisement entre ces mesures et la comparaison des taux des deux contextes (telle qu'affichée en fig. 12.2 *supra*), permet d'isoler les valeurs-*t* qui correspondent à une élision systématique et celles qui correspondent à une élision facultative.

Ainsi, sans surprise, les valeurs-*t* les plus élevées correspondent à des contextes où l'élision (devant #V) est presque systématique et où l'apocope (devant #C) n'a pas lieu : c'est le cas de l'article défini féminin *la* ($t = 532$; 97% de défauts devant #V), de l'article défini masculin singulier au cas régime *le* ($t = 350$; 99% __#V), du pronom personnel régime de première personne du singulier *me* ($t = 155$; 97% __#V) et de deuxième personne du singulier *te* ($t = 95$; 96% __#V) et des prépositions ($t = 86$; 87% __#V). Un seuil, dans la comparaison des valeurs-*t* et des taux contextuels d'apocope semble se dessiner avec le pronom personnel masculin régime *le* : sa valeur-*t* dépasse à peine 50 ($t = 58$), et il dispose de défauts non seulement devant initiale vocalique mais aussi devant initiale consonantique (99% __#V; 21% __#C). Néanmoins, il importe de considérer que ce pronom est particulier, dans la mesure où ses défauts devant initiale consonantique ne correspondent pas à de simples défauts mais à des cas d'enclise, le pronom *le* étant le plus productif pour les enclises avec un terme support précédent. De fait, si l'on peut considérer que l'élision l'emporte toujours sur l'enclise devant initiale vocalique (MARCELLO-NIZIA 2020c, GGHF, p. 685), la présence de défauts devant initiale consonantique, minoritaires, ne vient pas, ici, disqualifier l'élision. Ce seuil dans les valeurs-*t*, situé entre $t = 86$ (PRÉP) et $t = 58$ (PROPERs *le*), n'est donc pas effectif pour discréditer l'élision, mais cela tient uniquement à la spécificité de *le* pronominal, sujet également à l'enclise.

Néanmoins, aucune des catégories dont la valeur-*t* est inférieure à celle du pronom *le* n'a d'élision systématique. Même en l'absence de défauts devant initiale consonantique, l'élision demeure facultative pour les

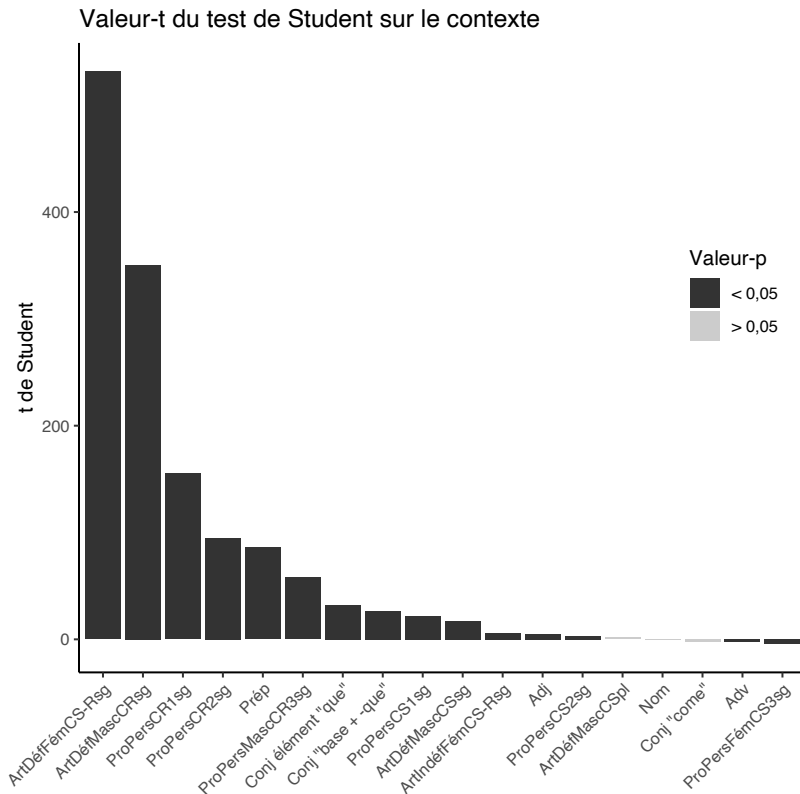


FIGURE 12.4 – Valeur- t du test de Student portant sur l'effet du contexte pour chaque catégorie.

conjonctions que sont l'élément *que* ($t = 31$; 41% __#V) et les locutions en *base + -que* ($t = 26$; 69% __#V), le pronom personnel sujet de première personne *je, jo* ($t = 22$; 48% __#V) et l'article défini masculin singulier sujet *li* ($t = 17$; 33% __#V). Les taux de défaut devant initiale vocalique passent en-deçà du seuil de 10% lorsque la valeur- t devient inférieure à 10, avec l'article indéfini féminin *une* ($t = 6$; 9% __#V; 1% __#C), les adjectifs féminins singuliers ($t = 4,84$; 6% __#V; 2% __#C) et le pronom personnel sujet de deuxième personne du singulier *tu, te* ($t = 3,1$; 1% __#V; 0% __#C). Les catégories dont la valeur- t est inférieure sont celles dont la valeur- p ne permet pas le rejet de l'hypothèse nulle (on ne peut démontrer un effet du contexte : article défini masculin CSpl *li*, noms, conjonction « come ») ou dont la valeur- t est négative (c'est le contexte préconsonantique qui génère le plus de défauts : adverbes et pronom personnel sujet féminin *ele*).

Ainsi, l'intersection entre les valeurs- t et les taux contextuels permet de définir la signification à donner à la valeur- t . Une valeur- t supérieure à 90 signifie que l'élision est systématique (96-99%) et non concurrencée par des défauts devant initiale consonantique (12.2 a). Une valeur- t supérieure à 80 signifie que l'élision est encore très forte et non concurrencée par des défauts devant initiale consonantique (87%; 12.2 b), tandis qu'une valeur- t inférieure à 60 peut signifier soit que l'élision est systématique mais concurrencée par les cas d'enclise (pronom *le*; 12.2 c) soit qu'elle n'est plus systématique (33-41%) mais n'est pas concurrencée par des défauts devant

initiale consonantique (12.2 d). En revanche, une valeur-*t* inférieure à 10 n'indique qu'une faible tendance à l'élision, éventuellement concurrencée par des cas de défaut devant initiale consonantique (12.2 e). Enfin, une valeur-*t* inférieure à 2 et accompagnée d'une valeur-*p* supérieure à 0,05 (seuil de significativité statistique) signifie que l'on ne peut affirmer que le contexte ait un effet significatif (12.2 f), tandis que les valeurs-*t* négatives dont la valeur-*p* est inférieure à 0,05 signifient que c'est le contexte préconsonantique qui génère le plus de défauts (12.2 g), mais cet effet est toujours faible (-2,5 pour les ADV et -4 pour le pronom personnel sujet *ele*). L'ensemble de ces observations est regroupé en (12.2), qui fournit une vision d'ensemble de ce que signifie la valeur-*t* de nos données vis-à-vis de l'élision.

(12.2) Force de l'élision selon les valeurs-*t*

	<i>t</i>	__#V	__#C	
(a) $t > 90$: élision systématique				
ARTDÉFFÉMSG <i>la</i>	532	97%	0%	
ARTDÉFMASC CRsg <i>le</i>	350	99%	0%	
PROPERS CR1sg <i>me</i>	155	97%	0%	
PROPERS CR2sg <i>te</i>	94	96%	0%	
(b) $90 > t > 80$: élision fortement dominante mais $< 90\%$				
PRÉP	86	87%	0%	
(c) $60 > t > 50$: élision systématique concurrencée par l'enclise				
PROPERS MASC CR3sg <i>le</i>	58	99%	21%	
(d) $40 > t > 10$: élision optionnelle				
CONJ élément <i>que</i>	32	41%	0%	
CONJ <i>base</i> + <i>-que</i>	26	69%	1%	
PROPERS CS1sg <i>je, jo</i>	22	48%	0%	
ARTDÉFMASC CSsg <i>li</i>	17	33%	0%	
(e) $10 > t > 2$: faible tendance à l'élision				
ARTINDÉFFÉMSG <i>une</i>	6	9%	1%	
ADJ FÉMSG	5	6%	2%	
PROPERS CS2sg <i>tu, te</i>	3	1%	0%	
(f) $t < 2$ & $p > 0,05$: absence d'élision (pas d'effet significatif du contexte)				
ARTDÉFMASC CSpl <i>li</i>	2	0,01%	0%	$p = 0,09$
NOM	0,5	0,01%	0,01%	$p = 0,6$
CONJ « come »	-2	62%	67%	$p = 0,06$
(g) $t < 0$ & $p < 0,05$: absence d'élision (effet significativement inverse du contexte)				
ADV	-2,5	68%	74%	
PROPERS FÉM CS3sg <i>ele</i>	-4	6%	14%	

12.2.2 Interprétation linguistique des valeurs-*t*, et intégration d'autres variables

La classification proposée en (12.2), si elle a l'intérêt de s'appuyer uniquement sur des valeurs discrètes, demeure imparfaite du point de vue analytique, puisqu'elle ignore des paramètres qui, ponctuellement, jouent un rôle prépondérant sur la motivation des valeurs en jeu. Ainsi, comme avancé précédemment, la présence forte des enclises pour le pronom personnel régime *le* (*a l, de l, ne l*; cf. sec. 9.2.2.1 et 9.2.3), si elle

fait chuter sa valeur-*t*, n'invalide en rien le fait que les défauts de ce pronom devant initiale vocalique sont bien des élisions, et que celles-ci sont systématiques. Elle ignore également la raréfaction progressive des enclises, du moins celles qui sont notées en deux mots dans le corpus.

Cette typologie ne parvient pas non plus à percevoir la diversité des situations au sein des catégories contenant plusieurs lemmes. Ainsi, pour les prépositions, elle ne représente pas le fait que la fréquence de la PRÉP « de » écrase les autres lemmes, dont les taux sont bien moins élevés : « entre » ne connaît que 68% de défauts devant initiale vocalique ($t = 23$; $p < 0,005$)² tandis que « contre » et « encontre » ne connaissent que 23% de défauts devant initiale vocalique ($t = 5$; $p < 0,005$).³

Par ailleurs, cette diversité est aveugle aux distributions diatopiques et diachroniques, du moins pour les catégories dont les taux d'élision sont inférieurs à 90%.⁴ Ainsi, elle ne dit rien de l'évolution du taux moyen d'élision du pronom personnel sujet *je*, *jo*, dont le taux synchronique de 48% masque une évolution de 0% d'élision graphique⁵ à ~60% à la fin du corpus. Cette typologie n'est pas non plus sensible à la distribution diatopique du déterminant défini masculin singulier *li*, dont l'élision est bien plus forte dans la moitié Sud-Ouest du domaine,⁶ à celle des défauts (dans les deux contextes) du pronom sujet *ele*, dont les défauts sont également plus présents au Sud-Ouest qu'ailleurs,⁷ ou à celle de l'article indéfini *une*, dont les défauts sont concentrés en Angleterre et en région Vendée, Deux-Sèvres.⁸ Enfin, pour les conjonctions, cette typologie est aveugle au fait que les formes défectueuses sont plus rares au centre du domaine (« come » __#C) ou sur une bande allant du Nord-Ouest au Sud-Est (*base* + *-que*, *que*, « come » __#V) que dans le reste du domaine.⁹

Plus généralement, elle est insensible à la situation de l'anglo-normand, qui se distingue pour bien des catégories. Cette spécificité de l'anglo-normand ne porte pas préjudice à la lecture globale pour certaines catégories, comme pour la PRÉP « de », dont le taux moyen d'élision est de 89%, le taux moyen a.n. de 59% et le taux moyen continental de 92%.¹⁰ Si, dans le cas de « de », l'anglo-normand ne vient pas biaiser trop lourdement les données continentales, il convient néanmoins de noter que l'élision systématique de cette préposition, si elle est valide sur le continent, ne l'est pas en anglo-normand. Dans d'autres cas, l'anglo-normand vient biaiser plus lourdement le taux de l'ensemble du domaine, comme c'est le cas pour les adjectifs féminins singuliers (devant #V, moy. a.n. : 11%; moy. a.fr.cont. 3%; moy. domaine : 6%)¹¹ ou pour le déterminant *une* (devant #V, moy. a.n. : 33%; moy. a.fr.cont. 18%; moy. domaine 9%; le taux continental est par ailleurs biaisé par le taux de la région Vendée, Deux-Sèvres, cf. *infra*).¹²

Enfin, cette typologie ne peut évidemment pas faire la part des choses entre les catégories pour lesquelles le corpus autorise une analyse fiable et celles pour lesquelles l'état des données biaise notre analyse. Ainsi, nous avons avancé en sec. 10.2.3 que le taux d'élision de 41% de l'élément *que* provient probablement du fait que cet 'élément' mélange pronoms et conjonctions pures, alors que le *-que* des locutions conjonctives en *base* + *-que* dispose lui d'un taux de 28% plus élevé (69%). Également à ce titre, cette typologie ne distingue pas les défauts qui tiennent à des collocations plus ou moins grammaticalisées. Ces défauts ont pourtant un rôle important dans certains des taux bas observés : ainsi, pour les PRÉP

2: Cf. fig. 6.17, et sec. 12.3.2.2 *infra*.

3: Cf. fig. 6.18, et *idem*.

4: L'effet des distributions diatopiques et diachroniques est nécessairement faible voire négligeable à partir d'un certain seuil de systématisme de l'élision.

5: Dans *Roland*, mais avec des élisions suggérées par la métrique, cf. sec. 9.1.1.1.

6: Cf. sec. 8.2.3.3 et (8.43) en sec. 8.3.3.

7: Cf. fig. 9.25.

8: Cf. sec. 8.2.3.3 et (8.43) en sec. 8.3.3.

9: Cf. sec. 10.2.4 et (10.26) en sec. 10.3.2.

10: Cf. sec. 6.3.5 et fig. 6.3.5.

11: Cf. (5.66) et (5.67) en sec. 5.3.5.

12: Cf. sec. 8.2.3.3 et (8.43) en sec. 8.3.3.

« entre », « contre » et « encontre », la majorité des formes défectueuses apparaissent devant le pronom personnel CR « eux »;¹³ il en va de même pour l'adverbe « ensemble » dans les occurrences *ensembl od* de Roland¹⁴ et pour une partie non négligeable des adjectifs féminins singuliers, comme développé en sec. 12.3.2 *infra*.

13: Cf. sec. 6.2.3.4 et (6.83) en sec. 6.3.1.

14: Cf. sec. 7.2.3.1 et (7.34) en sec. 7.3.1.

La modification de la typologie donnée en (12.2) *supra* par l'ensemble de ces éléments aboutit à la classification (12.3) en page suivante. Si celle-ci perd beaucoup en lisibilité du fait de la séparation de certains lemmes et de l'insertion des remarques diachroniques, diatopiques et liées aux collocations, elle représente la description la plus exhaustive de nos données. Les remarques contextuelles, diatopiques ou diachroniques sont précédées d'une flèche horizontale. Les catégories qui ont subi un déplacement vis-à-vis de la typologie précédente (12.2) sont indiquées par une flèche vertical avant l'entrée catégorielle; le motif du mouvement est expliqué, entre crochets, sous l'entrée. Une présentation graphique, moins détaillée mais permettant de saisir la vision d'ensemble, est donnée en fig. 12.6 après (12.3). Ces deux typologies sont nos classements finaux, ceux sur lesquels reposent les généralisations de ce chapitre et l'analyse du chapitre 14.

(12.3) Force de l'élision selon les valeurs-*t* et l'analyse linguistique

	<i>t</i>	__#V	__#C	
(a) Élision systématique				
↑ PROPERSMASC CR3sg <i>le</i>	58	99%	21%	
				[cat. promue par exclusion des cas d'enclise]
ARTDÉFMASC CRsg <i>le</i>	350	99%	0%	
ARTDÉFFÉMSG <i>la</i>	532	97%	0%	
PROPERS CR1sg <i>me</i>	155	97%	0%	
PROPERS CR2sg <i>te</i>	94	96%	0%	
↑ PRÉP « de »	86	89%	0%	
				→ 92% en a.fr.cont.; 59% en a.n.
				[cat. promue par dissociation des autres PRÉP]
(b) 40 > <i>t</i> > 10 : élision optionnelle				
(CONJ élément <i>que</i>	32	41%	0%	
				(→ 41% en a.fr.cont.; 24% en a.n.)
				(→ sur un axe N.-O. - S.-E.)
CONJ base + -que	26	69%	1%	
				→ 70% en a.fr.cont.; 26% en a.n.
				→ sur un axe N.-O. - S.-E.
↓ PRÉP « entre »	23	68%	0%	
				→ collocation avec le PROPERSCR « eux »
				[cat. rétrogradée par dissociation avec les autres PRÉP]
PROPERS CS1sg <i>je, jo</i>	22	48%	0%	
				→ croissance diachronique
ARTDÉFMASC CSsg <i>li</i>	17	33%	0%	
				→ élision surtout au S.-O.
(c) 10 > <i>t</i> > 2 : faible tendance à l'élision				
ARTINDÉFFÉMSG <i>une</i>	6	9%	1%	
				→ élision surtout en a.n. et Vendée, Deux-Sèvres
				→ 5% en a.fr.cont.; 33% en a.n. (__#V)
ADJFÉMSG	5	6%	2%	
				→ collocations <i>saint eglise, lung aleine, mal heure</i> etc.
				→ 3% en a.fr.cont; 11% en a.n. (__#V)
↓ PRÉP « (en)contre »	6	24%	0%	
				→ collocation avec le PROPERSCR « eux »
				[cat. rétrogradée par dissociation avec les autres PRÉP]
↑ ADV « ensemble »		3%	0%	
				→ collocation avec le parfait <i>od (Roland)</i>
				[cat. promue par dissociation avec les autres ADV]
PROPERS CS2sg <i>tu, te</i>	3	1%	0%	
(d) <i>t</i> < 2 & <i>p</i> > 0,05 : absence d'élision (pas d'effet significatif du contexte)				
ARTDÉFMASC CSpl <i>li</i>	2	0,01%	0%	<i>p</i> = 0,09
NOM	0,5	0,01%	0,01%	<i>p</i> = 0,6
CONJ « come »	-2	62%	67%	<i>p</i> = 0,06
				→ __{PROPERSCS, VERCjg} : <i>t</i> = -4,2 (<i>p</i> < 0,005); 68% __#V; 82% __#C
				→ moins de défauts au centre du domaine (__#C) et sur un axe N.-O. - S.-E. (__#V)
(e) <i>t</i> < 0 & <i>p</i> < 0,05 : absence d'élision (effet significativement inverse du contexte)				
ADV	-2,5	68%	74%	
PROPERSFÉM CS3sg <i>ele</i>	-4	6%	14%	
				→ défauts surtout au S.-O.

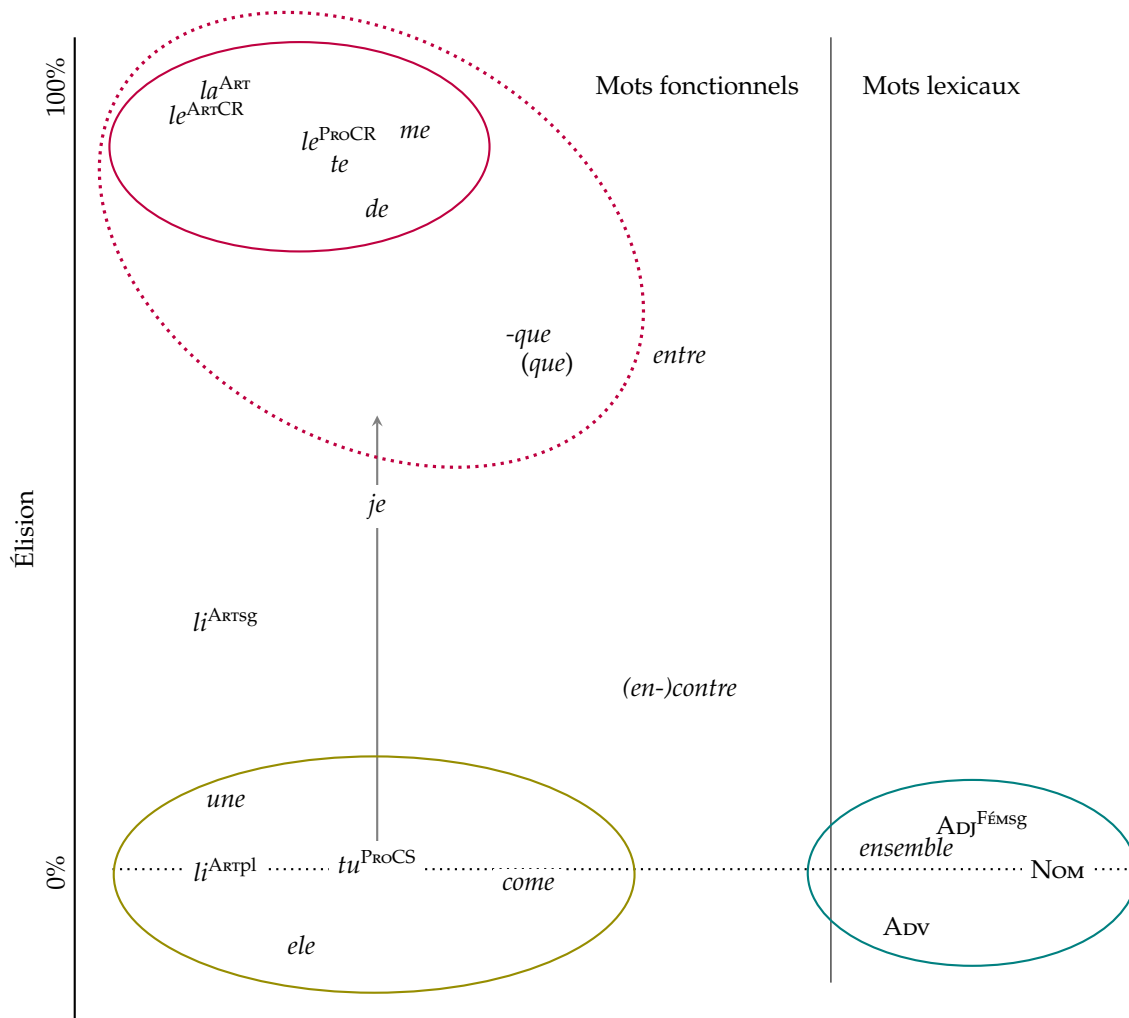


FIGURE 12.5 – Distribution des taux d'élision en fonction des des catégories et items. Un taux d'élision négatif signifie qu'on trouve plus de défauts devant #C que devant #V. Les valeurs des ordonnées sont approximatives.

Alternativement aux données quelque peu foisonnantes de la typologie en (12.3), nous proposons le graphe en fig. 12.6. Sur ce graphe, la position des catégories ou items sur l'axe des abscisses est arbitraire et illustre une logique d'alignement par catégorie, tandis que la position sur l'axe des ordonnées reflète approximativement le taux d'élision. Certaines des catégories de la typologie (12.3) sont reproduites par des ellipses sur le graphe. Ainsi, les items à élision systématique (12.3 a) sont regroupés dans une ellipse violette, à laquelle est adjoint *que* dans une ellipse en pointillés. Les items montrant une très faible tendance (12.3 c) à l'élision ou une absence d'élision (12.3 d - 12.3 e) sont regroupés dans deux ellipses, l'une ocre s'ils sont fonctionnels, l'autre cyan s'ils sont lexicaux. Le pronom personnel sujet *je* est indiqué avec une flèche, qui représente son évolution diachronique dans notre corpus ; l'étiquette « je » est placée au niveau du taux moyen d'élision du pronom.

Ce graphe permet donc une simplification analytique des types organisés selon la valeur *t* en (12.3). Il illustre également la séparation très claire entre

les mots fonctionnels et les mots lexicaux, avec uniquement *ensemble* (à cause des seules occurrences *ensembl od* de *Roland*) et les adjectifs féminins singuliers qui disposent de taux d'élision positifs, mais toujours très faibles. Quant aux adverbes et à *ele*, leur taux d'élision négatifs signifient que l'on trouve plus de défauts devant initiale consonantique que devant initiale vocalique, ce qui signifie que leurs défauts, même devant initiale vocalique, ne sauraient tenir à des élisions. De ce fait, ces catégories demeurent hors-champs de notre analyse, qui se concentre sur l'élision.

12.3 Généralisations morphosyntaxiques et phonologiques

La classification présentée en (12.3) n'est motivée par aucun critère morphosyntaxique ou phonologique. Elle révèle néanmoins deux faits de première importance, reflétés sur le graphe en fig. 12.6. D'une part, seuls les mots fonctionnels sont sujets à une élision systématique (12.3 a) ou optionnelle mais conséquente (12.3 b), tandis que les mots lexicaux ne connaissent qu'une faible élision (ADJ en 12.3 c), fortement dépendante de collocations et liée au dialecte anglo-normand, ou ne connaissent pas d'élision (NOM en 12.3 d ; ADV en 12.3 e). La réciproque de ce constat n'est en revanche pas vraie, dans la mesure où des mots fonctionnels comme l'article défini pluriel *li*, la conjonction « come » (12.3 d) ou le pronom sujet *ele* (12.3 e) ne connaissent pas d'élision.

(12.4) CONDITION MORPHOSYNTAXIQUE

Ne sont candidats à l'élision que les mots fonctionnels.

La seconde caractéristique remarquable tient à la taille des éléments impliqués : parmi les mots fonctionnels, tous ceux dont l'élision est systématique sont monosyllabiques de structure CV (c'est-à-dire sans consonne finale). Il en va de même pour ceux dont l'élision est optionnelle, à condition toutefois de considérer que les conjonctions en *base* + *-que* ne sont que des locutions et que *-que* est ici traité comme un mot, et que *entre* ne connaît de défauts que dans le cas de collocations avec le pronom personnel régime « eux ». Les seuls mots fonctionnels monosyllabiques non sujets à l'élision sont l'article défini masculin pluriel *li* et le pronom personnel sujet *tu*. Enfin, les mots lexicaux polysyllabiques autres que *entre* ne connaissent qu'une légère élision, également motivée par des collocations pour les prépositions « contre », « encontre », « ensemble », ou hors collocation mais avec une distribution diatopique fort inégale pour l'article *une*, voire ne connaissent pas d'élision en tant que telle, comme c'est le cas pour la conjonction « come » et le pronom sujet *ele* qui ont non seulement une distribution diatopique inégale mais également plus de défauts devant initiale consonantique que devant initiale vocalique.

(12.5) CONDITION DE STRUCTURE CV (v1)

Un mot fonctionnel n'est candidat à l'élision graphique que s'il se réduit à une structure *consonne* + *voyelle*.

Ainsi, une condition à l'élision semble être le fait que les mots fonctionnels soient monosyllabiques de structure CV. Cette condition, en

apparence non suffisante (cf. ARTDÉFMASCCSpl *li* et PROPERSCS2sg *tu*), semble néanmoins nécessaire : les mots fonctionnels polysyllabiques ne connaissent l'élision que marginalement ou dans le cas de constructions plus ou moins lexicalisées, la lexicalisation changeant le traitement qui leur échoit.

En marge de ces conditions morphosyntaxiques et phonologiques, une troisième condition semble jouer, autorisant des mots ne correspondant pas aux conditions (12.4) et (12.5) à perdre leur <e> dans certaines constructions impliquant un lien étroit avec le mot suivant. Cet aspect est abordé *infra*, dans une section dédiée aux collocations (sec. 12.3.2). Avant cela, la section 12.3.1 tente de fusionner provisoirement les conditions (12.4) et (12.5) en une seule condition phonologique, et développe les conséquences de cette fusion sur l'intégration des monosyllabes fonctionnels CV qui résistent à l'élision.

12.3.1 Élimination de la condition morphosyntaxique : COND CV

La condition (12.5) peut être étendue au-delà des mots fonctionnels, dans la mesure où, d'une part, l'élision ne concerne que les schwas dans les mots lexicaux, qui ne peuvent être la seule voyelle d'un mot lexical, et où d'autre part les mots lexicaux comportent une voyelle tonique - ou du moins susceptible de recevoir l'accent. On peut donc reformuler (12.5) en (14.1), ce qui élimine le besoin de recourir à la condition morphosyntaxique (12.4).

(12.6) COND CV

Un mot n'est candidat à l'élision graphique que s'il se réduit à une structure *consonne + voyelle atone*.

Ainsi, l'énoncé en (14.1) prédit efficacement le schisme entre, d'une part, les monosyllabes fonctionnels qui s'élident systématiquement (12.3 a) ou optionnellement (12.3 b) et d'autre part les mots polysyllabiques, fonctionnels ou lexicaux, qui ne connaissent l'élision que dans des collocations ou à des taux très faibles. Cet énoncé ne fait que délimiter l'ensemble qui est susceptible de subir l'élision, mais il ne prétend en rien que tous les items qui répondent à ces conditions sont effectivement atteints par l'élision. Or, un développement morphosyntaxique ou diachronique peut permettre de lui procurer cette puissance en évacuant les contre-exemples, faisant passer COND CV du statut de condition nécessaire à celui de condition suffisante. C'est le cas pour le pronom personnel sujet de première et de deuxième personne du singulier (sec. 12.3.1.1), ainsi que de l'article défini masculin pluriel *li* (sec. 12.3.1.2).

12.3.1.1 COND CV et les pronoms sujets de 1sg et 2sg

Dans le tableau du paradigme des pronoms personnels de l'ancien français que nous avons produit en fig. 9.1 (sec. 9.1.1), nous avons noté que les pronoms sujets de 1sg et 2sg commençaient par ignorer la division entre série forte et série faible, et que, lorsque cette distinction semblait apparaître au XII^e s., elle produisait néanmoins de nombreux

homographes entre les deux séries. Cette absence initiale de division puis l'intense homographie entre les deux séries peut permettre d'expliquer que ces formes ne s'élident pas, au début du corpus pour *je, jo*, et sur l'ensemble du corpus pour *tu, te*.

Je, jo. Pour le pronom de 1sg, seules les formes *gié* et *jou* n'appartiennent qu'à la série forte, tandis que toutes les formes de la série faible appartiennent aussi à la série forte. De même, pour le pronom de 2sg, la forme *tu* appartient aux deux séries, tandis que la forme *te* n'appartient qu'à la série faible. De ce fait, nos données concernant les formes faibles de ces pronoms contiennent toutes les formes fortes qui sont homographes d'une forme faible. Sur ce fondement, l'application de la règle COND CV rendue sensible à l'accent en (14.1) peut permettre de réhausser artificiellement le taux d'occurrences élidées du pronom de 1sg : en contexte syntaxique de forme forte, et peu importe la voyelle en question, le pronom n'est pas déclaré candidat à l'élision par l'effet de COND CV. Ainsi, en contexte syntaxique faible, *jo /ʒo/* ou *je /ʒə/* est déclaré apte à l'élision par COND CV dans la mesure où il a une structure CV atone, tandis qu'en contexte de forme forte, *je /ʒə/* ou *'ʒe/*,¹⁵ *jou /ʒu/* ou *gié /ʒje/* sont déclarés inaptes à l'élision par COND CV, ayant soit une structure CV qui n'est pas atone, soit une structure non CV.

De fait, la progression de l'élision graphique du pronom personne sujet de première personne, de 0% dans *Roland* à une moyenne de ~60% à la fin de notre corpus, procède non seulement du dédoublement d'une série unique en deux séries, l'une forte et l'autre faible, mais aussi de l'élargissement progressif des contextes syntaxiques où le pronom s'élide (MARCELLO-NIZIA 2015, p. 36-37). On peut tout à fait admettre que cet élargissement n'est pas celui du domaine de l'élision, mais du domaine de la série faible, et ce d'autant plus que toutes les formes non élidées faibles sont homographes d'une forme forte. Dans ce cas, la forme faible du pronom personnel sujet de première personne n'est pas dotée d'un taux d'élision anormalement bas de 48%, puisque les 52% manquants peuvent très bien correspondre à des contextes où c'est la forme forte qui est appelée par la syntaxe.

Tu, te. On ne peut en revanche tenir le même discours à propos du pronom personnel sujet de deuxième personne du singulier. Si MARCELLO-NIZIA (2020c, GGHF, p. 679-681) indique que les premières élisions graphiques de *tu, te* apparaissent au même moment que celles de *je*, ce pronom n'est élidé qu'extrêmement rarement dans le NCA : sur 1486 occurrences prévocales, il n'est élidé que dans 18 occurrences, dont la majorité sont des erreurs d'annotation où le pronom est régime et non sujet. L'élision de ce pronom est donc véritablement marginale dans notre corpus. Ce fait peut s'expliquer de deux manières différentes : soit il y a quelque chose dans ce pronom qui résiste à l'élision, soit l'expansion du domaine des formes faibles n'a pas véritablement lieu pour *tu* comme elle a lieu pour *je* et dans ce cas COND CV rejette naturellement la possibilité d'élider un pronom en contexte fort. Nous verrons dans la section suivante que la première hypothèse est tout à fait plausible, et évite la circularité qui consiste à définir le statut faible ou fort des positions syntaxiques par les formes et celui des formes par leurs positions syntaxiques. Enfin, comme noté en sec. 9.1.1.1, l'élision du pronom sujet

15: Schwa ne peut, en principe, être tonique. Néanmoins, il n'est pas sûr que la séparation entre formes fortes et formes faibles soit réductible à une séparation entre formes toniques et atones ; cf. sec. 14.3.1.1 sous 14.50.

MARCELLO-NIZIA (2015). « De JE à J' en français : étapes vers l'élision, interactions entre phonétique et syntaxe ».

MARCELLO-NIZIA (2020). « Les pronoms personnels, et EN, Y et ON ».

de 2sg n'est pas devenue systématique en français moderne, où elle est réservée à des contextes où l'expression est peu surveillée (*tu aimes ~ t'aimes*), ce qui plaide pour maintenir son déni de l'élision.

12.3.1.2 COND_{CV}, *li* sg élidable et *li* pl non élidable

Les articles masculins sujets singulier *li* et pluriel *li* sont homographes et il n'y a pas de raison de supposer qu'ils ne sont pas également homophones. Si l'on s'en tient à ce constat, il est impossible de prédire sur un fondement phonologique le fait que *li* sg peut être sujet à l'élision (certes modérée : 33% des occurrences prévocales) tandis que *li* pl n'est jamais sujet à l'élision. Or il est également difficile de prédire cette différence sur une base syntaxique, dans la mesure où les deux articles n'ont aucune raison de se trouver dans des contextes syntaxiques différents. On ne peut non plus utiliser la morphologie pour tenter de les distinguer, puisque si l'homophonie du singulier et du pluriel au même cas au masculin est certes inhabituelle, imaginer placer un suffixe de pluriel sur *li* pl reviendrait soit à postuler un suffixe jamais interprété (et donc difficile à transmettre, lors de l'acquisition de la langue, autrement que par l'absence d'élision)¹⁶ soit à placer un diacritique sur celui-ci, à la manière de ce que proposait par exemple la phonologie des années 80.¹⁷ La question est donc : comment faire en sorte que *li* pluriel soit déclaré inapte à l'élision par COND_{CV} tandis que *li* sg, comme les autres monosyllabes fonctionnels, est déclaré apte à l'élision ?

***Li* pl.** Une solution, ici, peut tenir à la phonétique historique. Il est d'ordinaire tenu pour acquis que les voyelles hautes /i/ et /u/, issues de *i* et *ū* latin, sont les seules à ne pas connaître de diphtongaison. Mais, comme le remarque STRAKA (1979, p. 233), celles-ci pourraient très bien avoir produit une diphtongue invisible, ce qui donnerait pour /i/ → /iɪ/ → /ij/ avant la monophthongaison en /i/. On peut, avec STRAKA (1979, p. 233) et SÉGÉRAL et SCHEER (2020, GGHF, p. 302), considérer que l'absence d'attestation d'une telle diphtongue « est ici simplement triviale » (GGHF, p. 302); admettre cette diphtongaison permet par ailleurs d'unifier le paradigme des diphtongaisons, qui concernent alors toutes les voyelles. Si l'on admet une étape *ILLĪ* → /lij/ pour *li* pluriel, cette étape peut faire échouer la candidature de l'article à l'élision : /lij/ n'est pas conforme à la structure CV demandée par la règle COND_{CV}, et est donc rejeté. Il s'agit, à notre connaissance, du seul cas de mot fonctionnel se terminant par /i/ ← *i* en a.fr.

***Li* sg.** Quant à l'article singulier *li*, sa formation n'est évidemment pas étymologique (*ILLE* ne peut évoluer en *li*) et procède selon toute évidence d'une analogie avec *ILLĪ* → *li* au pluriel. Il est nécessaire d'envisager que c'est cette analogie qui est responsable de la faculté du singulier à l'élider alors que le modèle sur lequel il est formé ne s'élide pas. Selon l'époque de cette analogie, deux scénarii se dessinent, mais aboutissent à la même situation — aussi le choix de l'un ou l'autre est indifférent pour notre argument. Si l'analogie a lieu alors que le modèle (*li* pluriel) dispose encore, dans sa forme sous-jacente mais pas dans sa forme de surface, du /j/ final (/lij/ → [li]), elle peut ne recopier que la forme de surface. On

16: Ce qui tendrait, non nécessairement mais potentiellement, à ramener ce suffixe au statut de diacritique.

17: Voir, p.ex., la spécification de certaines entrées lexicales pour résister à une règle phonologique dans DELL (1985, p. 238), qui précise que ce n'est que « faute de mieux » qu'il a recours à un tel mécanisme.

STRAKA (1979). *Les sons et les mots. Choix d'études de phonétique et de linguistique.*

obtient donc le singulier /li/ à partir du pluriel /lij/ → [li] et, alors que le pluriel /lij/, qui n'est pas de structure CV, n'a pas droit à l'élision par l'effet de la règle COND CV, le singulier /li/, conforme au gabarit CV, y a droit. En admettant, ce qui est crucial dans les deux scénarii, que la règle COND CV s'est appliquée en proto-ancien français et que ses conséquences sont ensuite figées — la règle décide, en proto-ancien français, qui aura droit à l'élision et qui n'y aura pas droit, et la décision est finale — on dérive alors la situation de l'ancien français, avec *li* singulier élidable et *li* pluriel non élidable.

Alternativement, si l'analogie a lieu après la simplification de la représentation sous-jacente du pluriel /lij/ en /li/, elle devrait produire un *li* sg qui n'aurait pas non plus droit à l'élision, du moins si elle recopie non seulement sa forme de surface mais aussi son diacritique d'élision. Mais, s'il y a analogie, c'est qu'il y a faiblesse dans l'inertie du système, et rien n'impose que l'analogie ne profite de cette faiblesse d'inertie que sur un seul plan. En face de *li* pluriel non élidable, les autres monosyllabes atones de type CV sont majoritairement élidables, et leur élision est souvent systématique (*le, la, me, te* en 12.3 a *supra*). Dans ce scénario, il conviendrait alors de considérer que si *li* sg a bien pris pour modèle mélodique *li* pl (déjà sous la forme /li/), il a en revanche pris pour modèle de la capacité à s'élider la majorité des monosyllabes CV atones. On obtient ainsi, comme dans le scénario précédent, la situation de l'ancien français où *li* pl ne peut s'élider puisque sa forme pré-littéraire /lij/ s'est vue refuser l'élision par la règle COND CV, mais ici *li* sg copie la possibilité de l'élision de la plupart des monosyllabes atones de forme CV. Dans ce scénario, le problème est donc résolu en considérant *li* sg comme une matrice contenant d'une part l'information mélodique et d'autre part la capacité à s'élider. Cette matrice et l'origine analogique de la valeur de ses attributs sont données en (12.7).

$$(12.7) \quad li \text{ sg} \left[\begin{array}{l} \text{Mélodie} = /li/ \\ \text{Élision} = \checkmark \end{array} \right] \begin{array}{l} \leftarrow li \text{ pl} /li/ \\ \leftarrow \text{ART } le, la, \text{ etc.} \end{array}$$

La matrice en 12.7 n'est pas une représentation phonologique, elle synthétise simplement les deux propriétés d'origine diverse que prévoit ce scénario; ce sont ces propriétés dont l'analyse phonologique proposée *infra* doit rendre compte sans tomber dans l'usage d'un diacritique. Le même type de matrice peut servir à représenter *li* pl et sa prohibition de l'élision, qui doit également être marquée comme un attribut différent du contenu mélodique en tant que tel (12.8).

$$(12.8) \quad li \text{ pl} \left[\begin{array}{l} \text{Mélodie} = /li/ \\ \text{Élision} = \times \end{array} \right] \begin{array}{l} \leftarrow /lij/ \\ \leftarrow /lij/ \times \text{COND CV} \end{array}$$

Retour à *Tu, te*. Nous avons relevé dans la section précédente que le pronom personnel sujet de deuxième personne du singulier ne s'élidait pas dans notre corpus. Or ce pronom remplit le même critère diachronique que *li* pl : il comporte, en latin, une voyelle haute et longue. Si l'on admet ici aussi la diphtongaison des voyelles hautes longues, on obtient $t\bar{u} \rightarrow /tu/ \rightarrow /tuus/ \rightarrow /tuw/ \rightarrow /tu/ \rightarrow /ty/$.¹⁸ Si /lij/ ne passe pas le test COND CV et n'acquiert donc pas la possibilité de s'élider, /tuw/ est dans la même situation, et le refus de l'élision établi par COND CV en proto-ancien français se maintient en a.fr. (12.9). Cela vaut également

18: La forme /tuus/ est proposée par SÉGÉRAL (SÉGÉRAL et SCHEER 2020, GGHF, p. 302).

pour la forme faible *te* du CS, où la dissociation entre la dimension mélodique et l'attribut de possibilité d'élision est évidente (12.10).¹⁹ En somme, si *li* pl et *tu* sg sont les seuls monosyllabes fonctionnels de l'a.fr. à échapper catégoriquement à l'élision, c'est parce qu'ils n'étaient pas des monosyllabes de structure CV au moment de l'application de COND CV. Ainsi, les seuls monosyllabes fonctionnels descendants de \bar{i} et \bar{u} résistent systématiquement, en a.fr., à l'élision.

$$(12.9) \quad tu \text{ CS} \begin{cases} \text{Mélodie} & = & /ty/ \\ \text{Élision} & = & \times \end{cases} \begin{matrix} \leftarrow /tuw/ \\ \leftarrow /tuw/ \times \text{COND CV} \end{matrix}$$

$$(12.10) \quad te \text{ CS} \begin{cases} \text{Mélodie} & = & /tə/ \\ \text{Élision} & = & \times \end{cases} \begin{matrix} \leftarrow /ty/ \text{ affaibli} \\ \leftarrow /tuw/ \times \text{COND CV} \end{matrix}$$

De fait, cette étymologie du déni d'élision de *tu* ne brise pas nécessairement le lien avec la situation de *je*. Nous avons remarqué dans la section précédente qu'il pouvait exister un parallèle entre la situation du pronom sujet de 1sg, dont MARCHELLO-NIZIA (2015) observe l'élargissement progressif des contextes syntaxiques d'élision, et l'absence d'élision de *tu* dans notre corpus. Si, en très ancien français, aucun de ces deux pronoms ne s'élide (ce qui peut être dû à la syntaxe), la rétention de l'absence totale d'élision de *tu* dans notre corpus appelle à ne pas se satisfaire de la syntaxe. De fait, si *tu* et *te* CS ont bien les attributs décrits en (12.9) et (12.10), ils peuvent connaître la même évolution syntaxique favorable à l'élision que *je* sans que celle-ci ne déclenche l'élision, en l'absence de l'attribut ✓ pour la variable de possibilité d'élision héritée du test COND CV.

12.3.1.3 En somme

Les seuls monosyllabes fonctionnels de gabarit CV en a.fr. à ne pas connaître l'élision, *li* pl et *tu* CS, peuvent avoir eu une forme non CV en proto-ancien français, ce qui les disqualifie pour l'élision. En effet, le statut élidable ou non d'un monosyllabe fonctionnel est décidé par le test COND CV, statut qui s'est joué en proto-a.fr. et été visiblement hérité tel quel par l'a.fr. ensuite. L'attribut hérité de COND CV n'a pas été adopté lors de l'analogie à l'origine de l'article CSsg *li*, qui a plutôt pris l'attribut des autres articles de gabarit CV.

La distribution des attributs d'élision héritée de la condition COND CV est représentée par de nouvelles ellipses sur le graphe en fig. 12.6, page suivante. Les deux items qui n'appartiennent à aucune ellipse, *entre* et *(en-)contre*, disposent également du statut COND CV = ×, et leurs cas de défaut de <e> tiennent à des collocations, ce que développe la section suivante.

12.3.2 Les collocations

Les sections 12.3.1.1 et 12.3.1.2 ont traité d'une catégorie d'exceptions à COND CV : les monosyllabes apparemment de gabarit CV qui ne sont pas soumis à l'élision. Il s'agit donc d'une exception au caractère suffisant de COND CV, et nous avons montré que cette exception pouvait n'être qu'apparente. Un autre type d'exception doit être traité ici : les infractions au caractère nécessaire de COND CV, c'est-à-dire les cas où un mot qui n'est

19: (12.10) fait l'impasse sur l'hypothèse d'une simple analogie avec le CR, dont la forme *te* s'élide. Une telle analogie procéderait selon la dissociation observée pour *li* sg en (12.7), adoptant la forme du CR mais conservant le statut d'élision du CS.

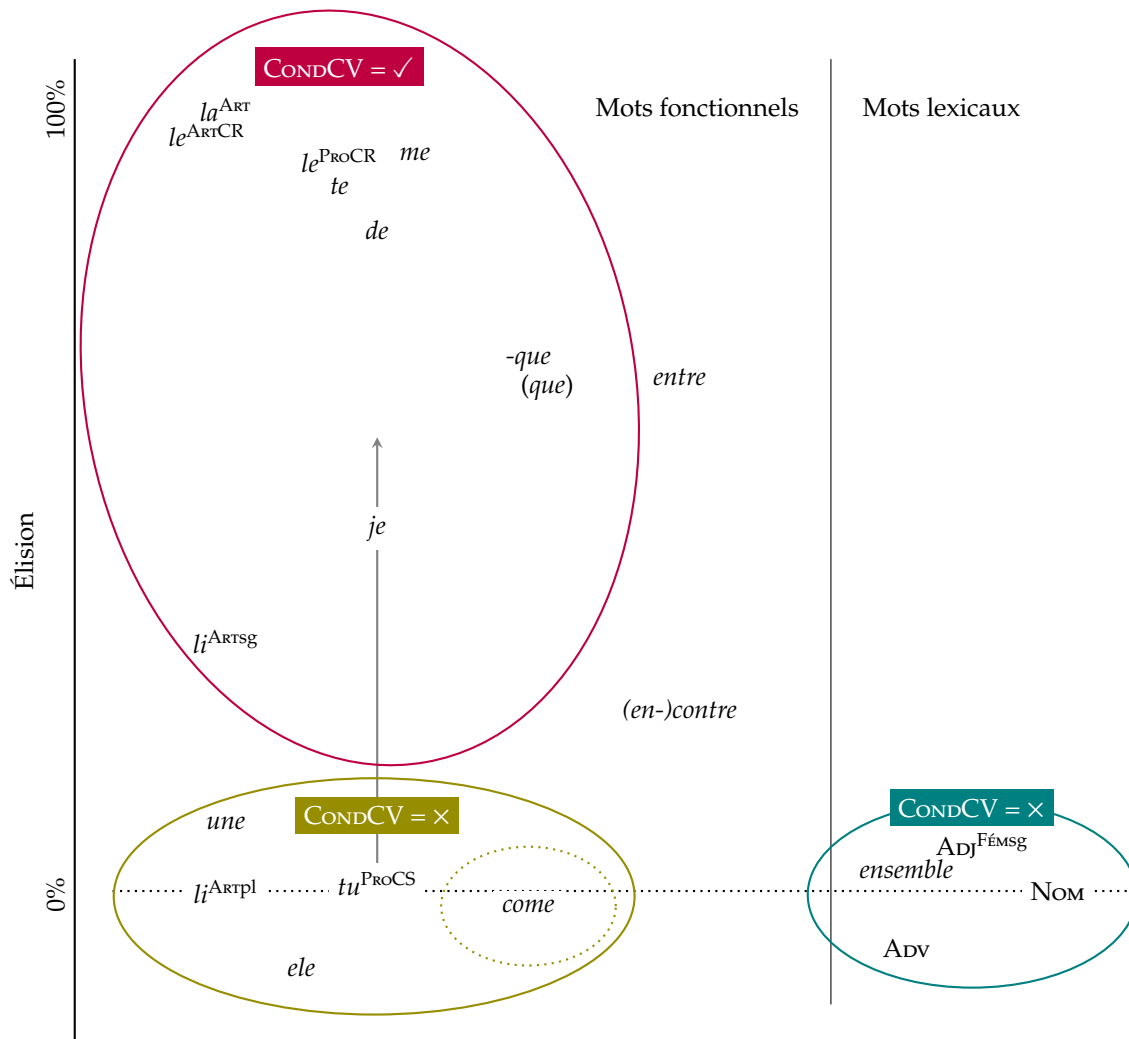


FIGURE 12.6 – Distribution des taux d'élision en fonction des de l'attribut de COND CV des catégories et items. Un taux d'élision négatif signifie qu'on trouve plus de défauts devant #C que devant #V. Les valeurs des ordonnées sont approximatives.

pas de gabarit CV est néanmoins sujet à l'élision. Dans nos données, cette exception concerne principalement les prépositions « entre », « contre » et « encontre », les adjectifs féminins singuliers et l'adverbe « ensemble ». Si les attestations de l'élision de ce dernier sont anecdotiques (*ensembl* n'est attesté que dans *Roland*), elles revêtent un caractère exemplaire du processus que l'on suppose à l'œuvre pour cette exception à la nécessité de COND CV.

12.3.2.1 *Ensembl od*, parangon de la collocation

Nous avons noté en sec. 7.2.3, 7.2.3.4 et 7.3.1 que le lemme « ensemble » apparaissait 17 fois sous la forme *ensembl* dans *Roland*. Ces 17 attestations sont non seulement toujours devant initiale vocalique, mais également toujours placées devant la préposition *od*. Parallèlement, le lemme « ensemble » est attesté six fois sans élision dans *Roland*, devant initiale vocalique (12.12) comme devant initiale consonantique (12.13), et aucune

de ces six occurrences ne précède immédiatement *od*. Il y a donc, ici, distribution complémentaire : *ensembl* précède *od*, et ailleurs seul *ensemble* existe.

- (12.11) *ensembl od lui rollant e oliver*
« avec lui sont Roland et Olivier »
(*Roland*, v. 104, trad. SHORT 1990)
- (12.12) *ensemble avum estét e anz e dis*
« Des jours, des ans nous avons été ensemble »
(*Roland*, v. 2028, trad. *id.*)
- (12.13) *e sunt ensemble plus de cinquante milie*
« ils sont en tout plus de cinquante mil »
(*Roland*, v. 1919, trad. *id.*)

L'absence de toute liberté paradigmatique et syntagmatique de la collocation *ensembl od* fait signe vers une grammaticalisation complète : la construction *ensembl od* ne fonctionne pas comme un *type* mais comme un *token*, c'est-à-dire comme une seule entrée lexicale, ses deux constituants ayant fusionné (BYBEE 2006). Mais cette fusion des constituants n'est pas suffisante pour générer l'éliision systématique : si la fusion résulte en /ensemblə.'od/, encore faut-il traiter le hiatus /ə.o/ formé sur le lieu de l'ancienne frontière entre les deux constituants ; or nous n'avons pas trouvé d'occurrence de hiatus réduits dans *Roland* (sec. 11.2). BYBEE (2000, 2002, 2006) et BYBEE et SCHEIBMAN (1999) proposent de considérer un effet intrinsèquement délétère de la fusion d'une construction en un *token* : le REDUCING EFFECT, qui aboutit à l'appauvrissement du matériel mélodique. Si la conjonction de ces deux effets peut expliquer la suppression du <e> dans *ensembl od*, nous serons néanmoins amenés à proposer une alternative dans notre analyse phonologique. Pour l'instant, nous acceptons comme provisoirement satisfaisante la conjonction de la fusion des constituants et de l'effet délétère du statut de *token*, et formalisons l'effet de ce type de collocation en (12.14).

- (12.14) COND_{COLL} (v1)
Un mot peut perdre son <e> final devant un second terme à initiale vocalique avec lequel il forme une collocation entièrement grammaticalisée.

12.3.2.2 Entr *eus* et (en)contr *eus*

La situation est légèrement différente pour les prépositions « entre », « contre » et « encontre », dont la liberté d'association avec l'élément suivant semble plus importante.

« **Entre** ». Pour « entre », on trouve en effet 393 formes *entr* devant *eus* (12.15), mais également 17 occurrences devant *eles* (12.16), 23 occurrences devant un verbe (12.17), 1 occurrence devant *autres* (12.18) et une devant *or* (le métal ; 12.19). Ainsi, la collocation n'est pas aussi stricte que pour *ensembl od*, cette dernière n'admettant aucune substitution ou insertion.

- (12.15) *lors parolent entr eus et dient*
« alors ils discutent entre eux, puis disent [...] »
(Lin, p. 3, *id.*)

BYBEE (2006). « From Usage to Grammar: The Mind's Response to Repetition ».

BYBEE (2000). « The phonology of the lexicon: Evidence from lexical diffusion ».

BYBEE (2002). « Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change ».

BYBEE (2006). « From Usage to Grammar: The Mind's Response to Repetition ».

BYBEE et SCHEIBMAN (1999). « The effect of usage on degrees of constituency: the Reduction of *don't* in English ».

Lin : ROBERT DE BORON (1213*). *Merlin*. Ms. A ; Dees : Yonne ; compo. Sud-Est, ca 1213 ; ms. bourg., ca 1250 ; CR2 (MICHÀ 1980) ; DEAF : MerLinM.

- (12.16) *entr eles s estoit mis*
« il s'était placé entre elles »
(chret1, p. 9, *id.*)
- (12.17) *enfanz qui s entr amerent*
« des jeunes gens qui s'aimaient l'un l'autre »
(deusamS, p. 1, trad. pers.)
- (12.18) *il ne se deigna [...] n a sei sozmetre ne ses deus mains entr autres metre*
« il ne deigna ni se soumettre ni placer ses deux mains dans celles de quelqu'un d'autre »
(benoit, p. 46, *id.*)
- (12.19) *vint somiers chargiés entr or fin et argent*
« vingt bêtes de somme chargées d'or fin et d'argent »
(flo, p. 6, *id.*)

Par ailleurs, cette collocation n'atteint pas le stade de la systématique. Si l'on s'en tient à la construction la plus attestée, *entre + PROPersCR3pl*, pour 410 graphies *entr eus* ou *entr eles*, on trouve également 137 occurrences *entre eus*. Même s'il n'est pas systématique, le défaut de cette construction demeure majoritaire, concernant 75% des occurrences « *entre* » + *PROPersCR3pl*. Il est donc nécessaire, si l'on veut autoriser « *entre* » à perdre son <e> devant *eles*, de modifier CONDCOLL (12.14) pour admettre des constructions disposant d'une liberté flexionnelle.

« (En)contre ». Les prépositions « *contre* » et « *encontre* » peuvent également être défectueuses lorsqu'elles sont associées à un pronom régime. Sur 13 occurrences défectueuses *contr*, 6 précèdent le *PROPersMASC CR3pl eus* (12.20). Le défaut dans cette construction n'est pas systématique, et est même minoritaire : aux 6 occurrences *contr eus* s'opposent 26 occurrences *contre eus* (12.21).

- (12.20) *eneas de bien loin les vit contr els ala*
« Énée les vit au loin ; il alla à leur rencontre »
(eneas, p. 13, trad. pers.)
- (12.21) *elle vint contre eulz*
« elle vint à leur rencontre »
(helc, p. 22, *id.*)

Trois occurrences défectueuses *contr* apparaissent dans des contextes où la collocation, si elle n'est guère attestée par les fréquences de notre corpus, est enregistrée comme complètement lexicalisée par certains dictionnaires : *contr aval* et *contr avaul* (floov, p. 15, 26) et *contr amont* (hista, p. 104) correspondent aux entrées « *cuntreval* » et « *contremont* » de l'AND et « *contraval* », « *contreval* » et « *contremont* » du GD. De fait, la préposition « *contre* » est souvent employée en préfixe (cf. sec. 6.2.3.4), ce qui accredit l'idée que CONDCOLL (12.14) doit exister sous une forme plus libre, admettant une lexicalisation partielle : *contre* peut fusionner avec de multiples seconds termes, devenant à terme un préfixe. La modification, ici, ne porte pas que sur la propriété de genre du second terme, mais sur une liberté de lemme de celui-ci.

En revanche, pour « *encontre* », l'association semble assez stricte : 6 des 7 occurrences défectueuses *encontr* précèdent le pronom personnel *eus* (12.22). L'occurrence restante voit cependant la préposition *a* s'interposer entre *encontr* et un autre pronom, le CR indirect singulier *lui* (12.23).

chret1 : CHRÉTIEN DE TROYES (1180*). *Le Chevalier au lion & Perceval*. Dees : Hainaut ; compo. champ. mérid., ca 1180 ; ms. champ., ca 1238 ; MS (transcr. inédite R. L. H. Lops) ; contenu présent ailleurs dans le corpus : incipit du *Chevalier au lion* (v. 1-998) dans yvf, et *Perceval* (v. 69-2000) dans perr.

deusamS : MARIE DE FRANCE (1165*). *Les Deus Amanz*. Ms. S ; Dees : nil ; compo. a.n., ca 1165 ; ms. frc., ca 1300 ; MS (transcr. inédite sans transcrituteur indiqué).

benoit : BENOIT DE SAINTE-MAURE (1174*). *Chronique des ducs de Normandie*. Ms. T ; Dees : Vendée, Deux-Sèvres ; compo. poit., ca 1174 ; ms. poit., ca 1190 ; MS1 (BENOIT DE SAINTE-MAURE 1951-1967).

flo : ANON. (1213*). *Florence de Rome*. Dees : Nièvre, Allier ; compo. pic., ca 1213 ; ms. Est, ca 1300 ; CR1 (WALLENSKÖLD 1907-1909) ; DEAF : FlorenceW.

eneas : ANON. (1160*). *Éneas*. Ms. A ; Dees : Haute-Marne ; compo. norm., ca 1160 ; ms. lorr., ca 1200 ; CR2 (SALVERDA DE GRAVE 1964-68).

helc : ANON. (1285*). *Le roman de Helcanus*. Dees : région parisienne ; compo. pic., ca 1285 ; ms. frc., ca 1290 ; CR2 (NIEDZIEMSKI 1966) ; DEAF : HelcanusN.

floov : ANON. (1190*). *Floovant, Chanson de Geste du XIII^e siècle*. Dees : Bourgogne ; compo. Sud-Est, ca 1190 ; ms. bourg., ca 1325 ; MS1 (ANDOLF 1965) ; DEAF : FloovA.

hista : WAUCHIER DE DENAIN (1213*). *Histoire ancienne jusqu'à César (Estoires Rogier)*. Ms. [BFN fr. 20125 & Vienne Nat. Bibl. 2576] ; Dees : [Haute-Marne] ; compo. Belgique, ca 1213 ; ms. Nord-Est, ca 1287 ; MS1 (WAUCHIER DE DENAIN 1995) ; DEAF : HistAncV.

- (12.22) *vos alez encontr eus*
 « vous allez à leur rencontre »
 (alia, p. 13, trad. pers.)
- (12.23) *est coruz encontr a lui*
 « [il] a couru à sa rencontre »
 (martin2, p. 12, id.)

La catégorie de collocation que signale l'occurrence *encontr a lui* n'est pas celle où la grammaticalisation aboutit à un *token* indivisible comme pour *ensembl od*, mais une association où non seulement les traits morphologiques peuvent varier (comme pour *entr eus* ~ *entr eles*) mais où l'insertion d'un autre mot entre les deux termes de la grammaticalisation est également admise. De ce fait, la construction grammaticalisée n'est pas *encontre + eus* mais *encontre + PROPERS*. Il faut donc apporter une autre modification à *COND COLL* : non seulement celle-ci admet la liberté paradigmatique des traits morphologiques du pronom, mais elle admet également la non-adjacence entre ses deux constituants. Cet exemple, pour extrême qu'il soit, permet d'illustrer le degré minimal de grammaticalisation de la collocation qu'attestent nos données. Nous reformulons donc *COND COLL* en (12.24) ; désormais, toute référence à *COND COLL* renvoie à cette version et non à la version en (12.14).

- (12.24) *COND COLL* (v2).
 Un mot peut perdre son <e> final devant un second terme à initiale vocalique avec lequel il forme une collocation, y compris si celle-ci n'est pas entièrement grammaticalisée.

12.3.2.3 Collocation des adjectifs

La version (12.24) de *COND COLL* permet également de rendre compte d'une partie des adjectifs féminins singuliers en défaut de <e>. Si nous n'avons pas été en mesure de détecter des collocations pour la majorité des occurrences adjectivales défectueuses, cela n'invalide pas le caractère nécessaire de *COND COLL*, dans la mesure où les défauts de <e> ne concernent qu'une faible minorité des adjectifs féminins singuliers. Pour 40 228 occurrences annotées selon ces propriétés morphologiques, 228 sont en défaut de <e>, soit 0,57%. Au sein même des 45 lemmes que nous avons retenus, ces défauts ne concernent que 2,43% des occurrences, et les variations de taux entre les lemmes ne sont probablement qu'un effet de biais de sélection dû à la fréquence des lemmes (cf. sec. 5.2.3.1).

L'association entre collocation et défaut de <e> des adjectifs féminins singuliers a été étudiée en sec. 5.2.3.2 ; nous nous en tenons ici à rappeler la teneur de l'argument, à l'illustrer et à constater la conformité de ces exemples à *COND COLL*. Les cas repérés concernent les lemmes « saint », « long » et « mal ». Pour le lemme « saint », 14 occurrences sont défectueuses devant initiale vocalique et, sur ces 14 occurrences, 12 correspondent au syntagme *sainte eglise* (12.25). Pour le lemme « long », 5 occurrences sont défectueuses, dont 3 composent le syntagme *longue haleine* (12.26), qui a encore le sens littéral d'un souffle long et est enregistré en tant qu'expression par certains dictionnaires. Enfin, pour « mal », 12 occurrences défectueuses sur 25 précèdent le substantif « heure » (12.27), formant la collocation à l'origine du mot *malheur*. Parmi les 13 autres

alia : ANON. (1185*). *Roman d'Alexandre*. Ms. A ; Dees : Vendée, Deux-Sèvres ; compo. pic. et occ. orient., ca 1185 ; ms. poit., ca 1225 ; MS (ANON. 1937) ; DEAF : AlexArsL.

martin2 : PÉAN GÂTINEAU (1125*). *Vie de saint Martin*. Dees : Vendée, Deux-Sèvres ; compo. tourr., ca 1125 ; ms. poit., ca 1290 ; MS (PÉAN GÂTINEAU 1896) ; DEAF : PeanGatS1.

occurrences de *mal* au féminin, nous avons également repéré d'autres collocations étant enregistrées comme des mots simples ou comme des expressions par les dictionnaires, comme *mal estrine* (« mauvais présage », 12.28) ou *malaventure* (12.29).

- (12.25) *il suffrit le sacrefice de sey pur dieu e seint eglise*
« il endura d'être sacrifié, pour Dieu et la sainte Église »
(edmund, p. 48, trad. pers.)
- (12.26) *un lung aleine un corn suna*
« il fit sonner un cor d'un souffle long »
(ipo, p. 81, trad. pers.)
- (12.27) *a mal ure fu nez*
« il naquit à un mauvais moment [= sous de mauvaises auspices] »
(sergbH, p. 4, trad. pers.)
- (12.28) *diex vous doint mal estrine quant si desordenement parlés*
« Dieu vous donne un mauvais présage quand vous parlez avec tant de confusion » (gar, p. 3, trad. pers.)
- (12.29) *lo fol croit si l en uient domaige es laz chiet de mal auenture*
« le fou croit, lorsqu'il est en péril, qu'il est tombé dans le piège d'une mauvaise aventure »
(yzop, p. 22, trad. pers.)

S'il n'est pas possible de délimiter avec précision la frontière entre les collocations qui fonctionnent comme telles et celles qui procèdent de la simple association syntagmatique, nous notons néanmoins que ces collocations concernent la moitié ou plus des occurrences des lemmes en question. Ces collocations, qu'elles soient totalement ou partiellement grammaticalisées, répondent donc à CONDCOLL, qui est responsable de leur défaut de <e>.

12.3.2.4 En somme.

Une autre condition que CONDCV semble autoriser les défauts de <e> : la formation avec le mot suivant d'une collocation. Celle-ci peut être strictement spécifiée (*ensembl od*) ou autoriser une certaine liberté paradigmatisée (*entr eus, entr eles*) voire syntagmatique (*encontr a lui*). Ces défauts liés à une collocation représentent une quantité non négligeable des défauts de certains adverbes, adjectifs et de certaines prépositions, ce qui démontre la pertinence de cette condition. Nous avons provisoirement accepté une étiologie de ces défauts reposant sur l'affaiblissement de la frontière lexicale entre les termes de la collocation et sur l'effet délétère du traitement conjoint des termes de la collocation. Si l'on pourrait envisager de considérer que tous les défauts qui dépendent de CONDCOLL ne sont pas des élisions et doivent être exclus de nos notre analyse, nous verrons au dernier chapitre (sec. 14.3.2.2) qu'ils peuvent aussi témoigner des premières traces de l'élargissement du domaine computationnel de l'a.fr. Dans cette hypothèse, ce serait *par* les collocations et préfixations que s'exprimerait cet élargissement en premier.

edmund : DENIS PIRAMUS (1195*). *La Vie Seint Edmund le Rei*. Dees : [Angleterre]; compo. a.n., ca 1195; ms. a.n., ca 1325; MS2 (DENIS PIRAMUS 1935); DEAF : EdmK.

ipo : HUE DE ROTELANDE (ca 1185). *Ipomedon*. Compo. a.n., ca 1185; ms. a.n., ca 1190; CR (transcription inédite).

sergbH : GUISCHART DE BEAULIU (ca 1190). *Sermon*. Compo. a.n., ca 1190; ms. a.n., ca 1250; CR1 (BEAULIU 1909), DEAF : GuischartG.

gar : ANON. (1275*). *Le Garçon et l'aveugle*. Dees : Aisne; compo. pic., ca 1275; ms. pic., ca 1275; CR2 (ANON. 1969 [1921]); DEAF : GarçAvR².

yzop : ANON. (1275*). *Ysopet de Lyon*. Dees : Franche-Comté; compo. Est, ca 1275; ms. frcomt., ca 1290; MS (ANON. 1882); DEAF : YsLyonF.

12.4 Conclusion

Dans ce chapitre, nous avons proposé une synthèse de certaines variables pertinentes pour la génération des défauts de <e> (sec. 12.1), ce qui nous a permis de dériver des typologies relatives aux taux absolus de défaut des catégories (12.1.1), à leur taux contextuels (sec. 12.1.2) ou à la force de l'effet de ce contexte tel que mesuré par le test de Student (sec. 12.1.3). Le croisement de ces typologies nous a permis de définir les catégories pour lesquelles les cas de défauts tiennent à l'élision et celles pour lesquelles ce n'est pas le processus d'élision qui est à l'œuvre (sec. 12.2), d'abord en ne tenant compte que des taux par catégorie (sec. 12.2.1), puis en incluant des effets diachroniques, diatopiques, lexicaux ou liés à des biais dans nos données (sec. 12.2.2). Cette démarche aboutit à la description synthétique (12.3), sur le fondement de laquelle nous avons pu établir des généralisations (sec. 12.3).

En première intention, la typologie adoptée appelle deux généralisations de nature différente : l'une est morphosyntaxique tandis que l'autre est phonologique. Sur le plan morphosyntaxique, nous avons remarqué que seuls les mots fonctionnels montraient des taux importants d'élision. Sur le plan phonologique, nous avons noté que ces taux importants ne concernaient que des monosyllabes de structure *consonne + voyelle*. Nous avons ensuite remarqué (sec. 12.3.1) que l'absence de mot lexical monosyllabique atone autorisait la fusion de ces deux conditions, sous une forme provisoire appelée COND CV. Celle-ci ne déclare comme candidats légitimes à l'élision que des mots de structure *consone + voyelle* atones, et, en tant que condition nécessaire mais non suffisante, produit les restrictions correspondant aux traits majeurs de la typologie utilisée (sec. 12.3.1).

Nous avons envisagé le caractère suffisant ou non de cette condition, et avons abouti à la conclusion que COND CV semblait être une condition suffisante. En effet, un examen des monosyllabes fonctionnels de type CV résistant à l'élision a montré que ceux-ci pouvaient échapper à COND CV pour des raisons syntaxiques (PROPERSCS1-2sg *jo, tu*; sec. 12.3.1.1) ou de phonétique historique (ARTDÉFMASCSP1 *li* et PROPERSCS2sg *tu*; 12.3.1.2), dans la mesure où un étymon non CV peut être reconstruit pour l'article défini pluriel *li* et pour le pronom personnel sujet *tu*. Cette reconstruction nous a amené à admettre que COND CV n'est pas une règle synchronique de l'ancien français, mais qu'elle s'est appliquée dans la préhistoire du français et que l'ancien français ne pratique l'élision graphique que pour les formes ayant validé leur capacité à s'élider : le fait qu'un item lexical pratique ou non l'élision est établi et figé en proto-ancien français, puis transmis tel quel à l'ancien français (12.30). Ce postulat, non content de rendre compte de l'absence d'élision de *li* pl et de *tu* sujet, permet également d'expliquer comment *li* sg peut s'élider alors que ce n'est pas le cas du pluriel avec lequel il partage pourtant l'intégralité de son matériel mélodique.

(12.30) La capacité d'un mot à s'élider dépend d'un test de conformité à un gabarit CV effectué en proto-a.fr. et non applicable en synchronie postérieure.

Enfin, nous avons examiné les contre-exemples où le mot élidé n'est pas de structure CV (sec. 12.3.2). Ceux-ci appartiennent majoritairement à des

collocations, et nous avons illustré le fonctionnement de ces collocations pour l'adverbe « ensemble » (*ensembl od*, sec. 12.3.2.1), les prépositions « contre » et « encontre » (*(en)contr eus*, sec. 12.3.2.2) et pour quelques adjectifs féminins singuliers (sec. 12.3.2.3). Nous avons formalisé la génération des défauts de <e> de ces formes par une condition de collocation, COND_{COLL}. Cette condition repose sur l'idée que les frontières de mots peuvent s'affaiblir dans le cadre de la grammaticalisation d'une locution, et nous avons admis provisoirement que ce processus pouvait également avoir un effet délétère sur le contenu mélodique des mots impliqués.

Il ressort donc de ce chapitre que les monosyllabes de type CV dont V n'est pas tonique sont sujets à l'élision graphique, et que celle-ci peut aussi, plus ou moins marginalement, atteindre le premier terme d'une collocation grammaticalisée. Ceci représente les grandes lignes de ce dont devra rendre compte notre analyse phonologique. Celle-ci devrait être capable de distinguer d'une part les monosyllabes atones (et donc fonctionnels) sujets à l'élision et ceux qui résistent à l'élision, et d'autre part de distinguer ces monosyllabes atones des polysyllabes et monosyllabes lexicaux. Cette distinction doit être inscrite dans la forme lexicale synchronique des mots en a.fr., mais elle est héritée d'un état plus ancien de la langue.

Dans ce chapitre, il n'a été question que d'élisions graphiques. Le chapitre suivant oppose aux élisions graphiques celles qui sont induites par la métrique et la musication, dont nous verrons que les cibles sont différentes.

Hormis des mentions ponctuelles de données métriques, les données présentées jusqu'ici concernent l'élision graphique, c'est-à-dire l'absence d'un graphème vocalique final devant un mot à initiale vocalique. Or, comme le développe ce chapitre, la métrique et la musication et, à partir du *xvi*^e s. les discours des grammairiens, témoignent d'une situation bien différente de la situation synthétisée dans le chapitre précédent.

Dans ce chapitre, nous abordons donc les données métriques, musicales et métalinguistiques. Dans l'introduction (sec. 13.1), nous donnons quelques points de repère qui permettent de situer le cadre théorique utilisé (sec. 13.1.1), abordons les notions de patron métrique, de grammaire d'association et d'extramétricalité du point de vue théorique (sec. 13.1.2) puis leur application aux productions médiévales françaises (sec. 13.1.3). En fin d'introduction (sec. 13.1.4), nous abordons notre échec à détecter des irrégularités métriques significatives avec le PAM, et en tirons la conclusion que l'absence de ces irrégularités est significative.

Nous abordons ensuite la question de l'élision métrique telle qu'elle se déploie dans les vers (sec. 13.2) et en musication (sec. 13.2.1.1), avant d'aborder l'intersection potentielle entre élision et coupe épique (13.2.2), pour en déduire le domaine qui est celui de l'élision dans les vers (sec. 13.2.3).

Nous constatons par la suite (sec. 13.3) que ces informations métriques sont cohérentes avec le discours des grammairiens du *xvi*^e s. (sec. 13.3.1), moyennant l'intervention d'une parophonologie métrique et l'association forcée des domaines computationnels avec des constituants métriques (sec. 13.3.2). Enfin, nous constatons l'écart entre les données graphiques résumées dans le chapitre précédent et ces données métriques (sec. 13.4), et proposons de fonder sur l'élargissement du domaine computationnel aux *xii*^e s. et *xiii*^e s. (sec. 13.4.1) une interpolation des données cachées sous une graphie désormais figée dans un système graphique insensible à certaines élisions (sec. 13.4.2).

13.1 Introduction

L'objet de cette section est d'introduire et d'explicitier les outils théoriques que nous utilisons pour aborder la métrique et la musication. Nous ne cherchons pas, ici, à les justifier, mais simplement à donner au lecteur le minimum d'informations nécessaires pour qu'il puisse juger de nos résultats. Nous commençons par introduire l'approche générative de la métrique (sec. 13.1.1), trois de ses outils fondamentaux pour notre analyse (les patrons métriques, la grammaire d'association et l'extramétricalité; sec. 13.1.2) et leur application à nos données médiévales (sec. 13.1.3), avant de terminer par les résultats paradoxaux que fournit le PAM sur l'évaluation des paramètres étudiés (sec. 13.1.4).

13.1.1 Métrique générative

Le cadre théorique que nous adoptons est celui de la métrique générative, selon le nom qui lui a été donné par BEAVER (1968). Cette approche, fondée sur le programme de JAKOBSON (1960, 1969 [1922]), propose de décrire les énoncés versifiés au sein de la théorie linguistique :

Poetics deals with problems of verbal structure, just as the analysis of painting is concerned with pictorial structure. Since linguistics is the global science of verbal structure, poetics may be regarded as an integral part of linguistics. (JAKOBSON 1960, p. 350).

Sur ce fondement, et en lien avec la conception du langage et des langues et le formalisme qui se développent dans le cadre de la linguistique générative, la métrique générative apparaît à partir des articles fondateurs de M. HALLE et S. J. KEYSER (1966, 1971) comme une tentative de décrire la génération des énoncés métriques bien formés, se distinguant du langage ordinaire mais appartenant néanmoins (a) à la faculté de langage humain et (b) à la grammaire de chaque langue. L'approche générative a également été adoptée pour l'analyse de la musication — l'association entre texte et musique — notamment à partir des travaux de LERDAHL et JACKENDOFF (1987) et J. HALLE et LERDAHL (1993). L'extension de la démarche générative de la métrique à la musique vocale ne tient pas qu'au fait que certains textes mis en musique sont versifiés, mais procède du constat que ces deux formes spécifiques de langage sont intimement liées :

As far as we know, the default state of metrical poetry is to be sung, or at least rhythmically chanted, rather than spoken. Homer, the Serbo-Croatian guslars, rap artists, and the singers of Appalachian folk ballads are all exponents of poetry in what is arguably its most natural form. (HAYES et KAUN 1996, p. 244)

Que ce soit pour la métrique ou pour la musication, leur étude générative tente de se satisfaire des outils et concepts utiles à la description des langues, sans inventer de dispositifs *ad-hoc*. Les représentations et mécanismes fondamentaux de la métrique générative sont donc motivés indépendamment de la métrique, et la théorie métrique suit l'évolution de la théorie linguistique (BLUMENFELD 2016, p. 418). Notre propos n'est pas, ici, de proposer une théorie générale de la métrique médiévale, mais uniquement d'exposer les outils que nous utilisons pour aborder les vers et la musique vocale médiévale en tant que sources linguistiques.

13.1.2 Patron métrique, grammaire d'association et extramétricalité

Le vers est organisé en fonction d'un modèle abstrait appelé *mètre*. Ce mètre est constitué d'un ensemble de positions, dont le nombre, l'ordre et le statut de prééminence sont déterminés. Une *grammaire d'association* régit les différentes modalités par lesquelles les unités d'un énoncé linguistique peuvent être associées aux unités du mètre. Lorsque la grammaire d'association est violée, le vers est considéré mal formé ; lorsque la grammaire d'association est respectée, le vers est considéré

BEAVER (1968). « Progress and Problems in Generative Metrics ».

JAKOBSON (1960)). « Linguistics and poetics ».

— 1969 [1922]). *O československému slovesnému umění v srovnání s ruským*.

M. HALLE et S. J. KEYSER (1966). « Chaucer and the study of prosody »

— (1971). « Illustration and defense of a theory of the iambic pentameter ».

LERDAHL et al. (1987). *A Generative Theory of Tonal Music*.

J. HALLE et al. (1993). « A Generative Text-setting Model ».

HAYES et al. (1996). « The role of phonological phrasing in sung and chanted verse ».

BLUMENFELD (2016). « Generative Metrics : An Overview ».

bien formé. Par exemple, dans le cas du pentamètre iambique anglais, le mètre est constitué de cinq iambes, et l'association optimale entre mètre et énoncé linguistique aligne positions et syllabes dans un rapport de bijection, en respectant un alignement entre positions proéminentes (s pour *strong*) et syllabes accentuées et entre positions faibles (w pour *weak*) et syllabes atones, et aligne également les frontières de pieds avec des frontières de mots, comme dans le vers de Shakespeare en (13.1).

(13.1) (W S) (W S) (W S) (W S) (W S)
 | | | | | | | | | |
Of hand, of foot, of lip, of eye, of brow
 (SHAKESPEARE, sonnet 106, adapté de KIPARSKY 1977, p. 194)

Dans le vers de Shakespeare en (13.1), la congruence, c'est-à-dire l'association parfaitement bijective entre constituants linguistiques et constituants métriques, demeure imparfaite : certes les frontières de pieds sont associées à des frontières de mots, mais seulement la moitié des frontières de mots sont associées à des frontières de pieds. De fait, étant donné le patron métrique du pentamètre iambique anglais, la grammaire d'association ne peut exiger une congruence totale, le nombre de propositions syntaxiques pouvant être composées de cinq mots iambiques étant fort limité. BLUMENFELD (2016) propose l'exemple (13.2) comme illustration d'une congruence parfaite.

(13.2) (W S) (W S) (W S) (W S) (W S)
 | | | | | | | | | |
A- las, Ma- rie a- voids ma- rooned ca- noes
 (Adapté de BLUMENFELD 2016, p. 421)

Une grammaire d'association entre mètre et énoncé linguistique peut donc admettre un certain nombre de libertés, mais d'une manière toujours contrôlée : les associations entre deux éléments non équivalents (en. *mismatch*) qui sont permises au sein d'une tradition métrique sont licenciées dans la grammaire de cette tradition. Parmi les modalités d'association non congruente, une tradition peut autoriser un décalage entre frontières métriques et frontières linguistiques, et/ou autoriser un décalage entre la proéminence des syllabes et celle des positions. L'une des formes que peut prendre la non congruence de proéminence est illustrée par l'exemple de Milton en (13.3), où le mot *future* est de structure sw, mais associé à un pied ws. La non congruence entre frontières de pieds et frontières de mots est illustré par l'exemple de Robert Frost en (13.4), dont les deux derniers pieds sont associés à trois mots.

(13.3) (W S) (W S) (W S) (W S) (W S)
 | | | | | | | | | |
Be- yond all past ex- amp- le and fu- ture
 (MILTON, *Paradise Lost*, 10.840, adapté de KIPARSKY 1977, p. 201)

(13.4) (W S) (W S) (W S) (W S) (W S)
 | | | | | | | | | |
It shall not make a- gain all win- ter long

(FROST, « Onset », v. 4, adapté de BLUMENFELD 2016, p. 415)

Un dernier type de non congruence est important pour notre approche des vers de l'ancien français : l'extramétricalité. Dans ce type d'association, une syllabe n'est associée à aucune position métrique, ce qui en anglais peut advenir devant une frontière syntaxique et après une position forte.¹ C'est le cas dans le vers en (13.5), où la cinquième et douzième syllabes ne sont associées à aucune position métrique. Pour signaler que la non association d'une syllabe avec une position est licite dans une grammaire d'association donnée (c'est-à-dire est un cas d'extramétricalité), nous notons une telle syllabe entre chevrons '<' et '>'.¹

(13.5) (W S) (W S) (W S) (W S) (W S) (W S)
 | | | | | | | | | | | |
 That is the mad- <man.> The lov- er, all as fran- <tic>
 (SHAKESPEARE, *A Midsummer Night's Dream*, V, 1, v. 10, adapté de
 KIPARSKY 1977, p. 231)

13.1.3 Application à l'ancien français

Dans cette section, nous développons l'application que nous faisons des patrons métriques (sec. 13.1.3.1), de la grammaire d'association et de l'extramétricalité (sec. 13.1.3.2 et 13.1.3.3) aux données médiévales. Hormis pour la musication en sec. 13.1.3.3, l'exposé est essentiellement théorique, l'interprétation pré-théorique des données en question étant relativement consensuelle.

13.1.3.1 Patrons métriques

Les gabarits métriques que nous utilisons dérivent de ceux proposés pour le vers français classique par VERLUYTEN (1988), AROU (2019) et DELL et BENINI (2021). Contrairement au gabarit de l'alexandrin de Verluysen, nous ne plaçons pas de contraste de proéminence entre les positions avant la dernière position de chaque constituant métrique et, contrairement aux exemples du pentamètre iambique *supra*, nous ne recourons pas aux pieds.² De ce fait, les patrons métriques que nous utilisons sont composés de positions × dont la proéminence n'est pas spécifiée et de positions '× proéminentes en fin de constituant. Les patrons métriques des octosyllabes (8s), décasyllabes à césure 4/6 (10s) et alexandrins (12s) que nous utilisons sont donnés en (13.6 - 13.8); ce ne sont pas les seuls que nous utilisons, mais ce sont les plus communs.³

(13.6) 8s [× × × × × × × '×]_v

(13.7) 10s [[× × × '×]_h [× × × × × '×]_h]_v

(13.8) 12s [[× × × × × '×]_h [× × × × × '×]_h]_v

1: De fait, ces critères sont communs avec ceux de la coupe épique (cf. *infra*) en ancien français, quoiqu'ils autorisent une plus grande liberté dans le placement précis de la coupe.

VERLUYTEN (1988). « Introduction ».

AROU (2019). « Césure et parophonologie en français, en grec et en latin ».

DELL et al. (2021). *La Concordance chez Racine. Rapports entre structure grammaticale et structure métrique dans le théâtre de Racine*.

2: Hors de la littérature générative, on trouve de multiples références aux pieds en métrique française, y compris pour le vers classique; ce que désignent ces pieds correspond le plus souvent au rythme et non au mètre du vers, c'est-à-dire à sa prosodie linguistique et non à sa conformation à un patron métrique. S'il a bien existé une structuration rythmique interne à l'hémistiche ou au vers en ancien français, particulièrement marquée dans les textes les plus anciens et de moins en moins perceptible ensuite (NOYER 2002; RAINSFORD 2010, 2011b), nous ne postulons pas que celle-ci ait jamais eu une régularité suffisante pour être en mesure de contrôler la bonne formation du vers. Noyer (p. 159) note à ce propos : « no text at any period conforms absolutely to the Iambic Pattern in the sense of classical English verse (Shakespeare, Milton, or even Shelley) ».

3: *Aucassin et Nicolette*, par exemple, recourt à des patrons de 7 positions, et de 4 positions pour les vers orphelins.

Dans les représentations en (13.6 - 13.8), frontières de domaine et positions proéminentes sont redondantes : le constituant []_h (hémistiche) est dérivable à partir des positions proéminentes '×, tandis que le constituant []_v (vers) est inutile si l'on considère que chaque itération du patron correspond au maximum de linéarité possible. Si nous les indiquons ponctuellement lorsque cela éclaire un exemple, nous considérons que la représentation d'un patron métrique ne devrait pas comporter une telle redondance, et que les patrons *supra* sont représentés d'une manière plus optimale en (13.9 - 13.11).

(13.9) 8s × × × × × × × '×

(13.10) 10s × × × '× × × × × × '×

(13.11) 12s × × × × × '× × × × × × '×

En (13.12 - 13.14), ces patrons sont associés avec le premier vers de l'*Érec et Énide* de Chrétien de Troyes (Erecku), le premier vers de la neuvième laisse de *Roland* (roLand) et le premier vers de la cinquième strophe de la *Vie de saint Thomas Becket* de Guernes de Pont-Saint Maxence (becket).

(13.12) × × × × × × × '×

| | | | | | | |
Li vi- lains dit an son res- pit

« Le proverbe du vilain nous enseigne »
(*Érec*, v. 1, trad. DEMBOWSKI in CHRÉTIEN DE TROYES 1994)

(13.13) × × × '× × × × × × '×

| | | | | | | | | |
Blan- can- drins ad tut pre- me- reins par- led

« Tout le premier, Blancandrin prit la parole »
(*Roland*, v. 122, trad. SHORT 1990)

(13.14) × × × × × '× × × × × × '×

| | | | | | | | | | | |
Sei- gnurs pur a- mur Deu e pur sal- va- ti- un

« Seigneur, pour l'amour de Dieu et pour votre rédemption, »
(*Becket*, v. 21, trad. pers.)

13.1.3.2 Grammaire d'association et extramétricalité

Les exemples en (13.12 - 13.14) ont été sélectionnés pour illustrer l'association la plus simple - et la plus courante - entre syllabes et positions métriques. Il y a bijection : chaque position est associée à une syllabe, chaque syllabe est associée à une position. Par ailleurs, les positions proéminentes '× sont associées à une syllabe accentuable — nous utilisons le terme *accentuable* pour ne pas préjuger de l'accentuation concrète du vers, reprenant l'usage de RAINSFORD (2011a, p. 270-272) qui définit des *unités accentuelles possibles*. Du fait de la sous-spécification de proéminence des positions internes, nous ne relevons pas de discongruence

Erecku : CHRÉTIEN DE TROYES (1170*). *Erec et Enide*. Compo. ca 1170, ms. ca 1235, dial. champ. (CHRÉTIEN DE TROYES 2009b).

roLand : ANON. (ca 1100). *Chanson de Roland*. Compo. ca 1100, ms. ca 1120, dial. norm. (MOIGNET 1972a).

becket : GUERNES DE PONT-SAINTE-MAXENCE (ca 1173*). *Vie de saint Thomas Becket*. Compo. entre 1172 et 1174, ms. déb. XIII^e s., dial. Ouest (GUERNES DE PONT-SAINTE-MAXENCE 1936).

RAINSFORD (2011). « Dividing lines: the changing syntax and prosody of the mid-line break in medieval French octosyllabic verse ».

entre syllabes toniques et positions non proéminentes. De fait, ces vers semblent répondre à une grammaire d'association contenant les règles (13.15) et (13.16).

(13.15) $B_{1j}E_{1c}$.

L'association entre syllabes et positions se fait selon un rapport bijectif.

(13.16) $'\times \implies '\sigma$

Une position métrique proéminente est associée avec une syllabe accentuable.

Ces deux règles sont trop simplistes, même si l'on se restreint aux trois textes cités. En effet, le rapport de bijection connaît une infraction courante qui n'entrave en rien la bonne formation du vers : l'extramétricalité après une position proéminente (c'est-à-dire en fin d'hémistiche et de vers). Cette extramétricalité est responsable des coupes épiques et des terminaisons féminines (rimes féminines), comme illustré, pour ces trois textes, en (13.17 - 13.19). Comme en (13.5) *supra*, une syllabe extramétrique est notée entre chevrons.

(13.17) $\times \quad \times \quad \times \quad \times \quad \times \quad \times \quad \times \quad '\times$
 | | | | | | | |
qui mout valt mialz que l'an ne cui- <de>

« [chose qu'on dédaigne] vaut mieux qu'on ne le pense. »
 (Érec, v. 3, trad. DEMBOWSKI in CHRÉTIEN DE TROYES 1994)

(13.18) $\times \quad \times \quad \times \quad '\times \quad \quad \quad \times \quad \times \quad \times \quad \times \quad \times \quad '\times$
 | | | | | | | | | |
Li em- pe- re- <res> Car- les de Fran- ce dul- <ce>

« Charles, l'empereur de France la douce »
 (Roland, v. 16, trad. SHORT 1990)

(13.19) $\times \quad \times \quad \times \quad \times \quad \times \quad '\times \quad \quad \quad \times \quad \times \quad \times \quad \times \quad \times \quad '\times$
 | | | | | | | | | | | |
Se Je- su Crist le suef- <fre> ki de nus tuz est si- <re>

« Si Jésus-Christ, qui est notre seigneur à tous, le permet »
 (Becket, v. 12, trad. pers.)

La coupe épique n'est pas admise par tous les textes, et lorsqu'un texte la pratique, elle est minoritaire, mais pas marginale. Dans *Roland*, par exemple, 853 vers sont des décasyllabes bien formés contenant une coupe épique, ce qui représente 21% des vers du texte (PREMAT et POGGIO [à paraître]). Cette minorité n'est pas nécessairement due à une complexité supérieure de la coupe épique qui en réduirait la fréquence; elle peut, plus simplement, provenir du fait que la majorité des mots du lexique est composée d'oxytons. Le nombre de vers paroxytons est plus élevé (approximativement 1350, soit 34% du texte),⁴ mais du fait que le statut oxytonique ou paroxytonique est homogène au sein d'une laisse donnée (ce qui n'est pas le cas pour les coupes épiques), ce nombre peut être biaisé par les attendus superstructurels.

Il est donc nécessaire de réécrire $B_{1j}E_{1c}$ (13.15) *supra*, en adoptant une directionnalité analogue à celle que nous avons adoptée pour $'\times \implies$

4: Ce chiffre est une estimation. Nos données concernant *Roland* indiquent que le PAM a considéré 1353 vers comme étant féminins, mais nous n'avons pas entrepris la vérification manuelle de ces 1353 vers.

' σ (13.16). Plutôt que de décrire l'extramétricalité elle-même, nous l'admettons en négatif, en décrivant les règles d'associations directionnelles qui concernent les positions atones (13.20) et toniques (13.21), les syllabes atones qui ne sont pas la finale d'un paroxyton dont la tonique est associée à une position métrique proéminente (13.22), et les syllabes toniques (13.23). La condition contextuelle en (13.22) peut certes sembler peu élégante, mais elle permet de dériver l'extramétricalité naturellement : les paroxytons dont la syllabe tonique est associée à une position proéminente n'associent tout simplement pas leur posttonique.⁵

(13.20) $\times \implies \sigma$
Toute position métrique non proéminente doit être associée à une (et une seule) syllabe.

(13.21) ' $\times \implies '\sigma$
Une position métrique proéminente est associée avec une syllabe accentuable.

(13.22) $\sigma \implies \times \quad \neg / \left[\dots \begin{array}{c} '\times \\ \Downarrow \\ '\sigma \quad _ \end{array} \right]$

Toute syllabe atone doit être associée à une (et une seule) position métrique, sauf (\neg) si elle est la posttonique d'un paroxyton dont la tonique est associée à une position proéminente (= sauf lorsqu'elle se trouve immédiatement après une syllabe accentuable associée à une position proéminente et qui appartient au même mot).

(13.23) ' $\sigma \implies \{\times, '\times\}$
Toute syllabe accentuable doit être associée à une position, que celle-ci soit proéminente ou non.

Les règles d'association en (13.20 - 13.23) ne disent rien de l'élision, qui fait l'objet de la section suivante. En acceptant provisoirement l'anticipation de la règle de l'élision métrique, cette grammaire d'association rend compte des vers de la plupart des textes que nous avons étudiés avec le PAM. Le programme détecte ainsi plus de 94% de vers conformes au modèle de l'octosyllabe (13.9) dans 31 textes,⁶ et, dans la plupart des cas, les quelques pourcents manquants sont dus à des défauts de paramétrage du programme face à des formes graphiques rares. Dans le cas de *Roland*, en acceptant les coupes épiques et des élisions de monosyllabes fonctionnels non transcrits (*ço est* \rightarrow *c est*; *jo ai* \rightarrow *j ai*; cf. sec. 9.1.1.1 et *infra*), nous parvenons également à 93% de vers conformes au patron décasyllabique. La situation est en revanche moins évidente pour les textes anglo-normands, et ce peu importe leur mètre; pour *Roland*, le PAM ne parvient pas à dépasser les 91% de vers conformes au patron du décasyllabe en (13.11). Les 9% de vers non conformes au patron *via* la grammaire d'association ont été vérifiés manuellement et leur statut hyper- ou hypométrique a été confirmé. Ils peuvent se comprendre soit comme une corruption de l'édition ou de la source, soit comme des représentants de l'isosyllabisme approximatif de l'a.n., où les irrégularités métriques *a priori* non réductibles par une grammaire d'association ne sont pas rares et sont acceptées telles quelles par la bibliographie (SHORT 2007, p. 40).

5: Nous ne rentrons pas ici dans l'organisation interne de la grammaire d'association. Les règles (13.20) et (13.21) peuvent être envisagées comme des sous-catégories ou des modalités d'une règle plus générale; même chose pour (13.22) et (13.23). Nous proposons simplement des règles qui génèrent les configurations attestées et n'en génèrent pas d'autres.

6: 8s conformes à (13.20) - (13.23) : atrper, belinc, BenDuc1b, beroul, bestam, brut2, CharretteKu, CligesKu, desire, doon, enea1 et -2, eracle, ErecKu, espine, graelent, guingamor, lecheor, MirNdChartr, nabaret, PercevalKu, rosel, rosem1, -2 et -3, thebes2, trot, tydorel, tyolet, vergy, YvainKu.

13.1.3.3 Extramétricalité et musication

Comme exposé en sec. 2.3, nous recourons à des sources musicales en complément des sources métriques. Pour l'analyse de la musication, nous utilisons un modèle tripartite assez simple, développé notamment dans nos publications,⁷ et qui s'inspire de l'approche modulaire de KIPARKSY (2006) et de paramètres avancés notamment par DELL et J. HALLE (2009) et DELL (2015). Dans cette approche, non seulement le composant linguistique (le 'texte') est associé au mètre, mais il est aussi associé à la musique, tandis que la musique et le mètre sont également associés entre eux. L'ensemble, plus difficile à schématiser, peut être représenté, en trois dimensions, sous la forme d'un prisme triangulaire tel qu'en fig. 13.1; nous n'utiliserons pas de tels schémas, de lecture difficile, dans l'exposé.

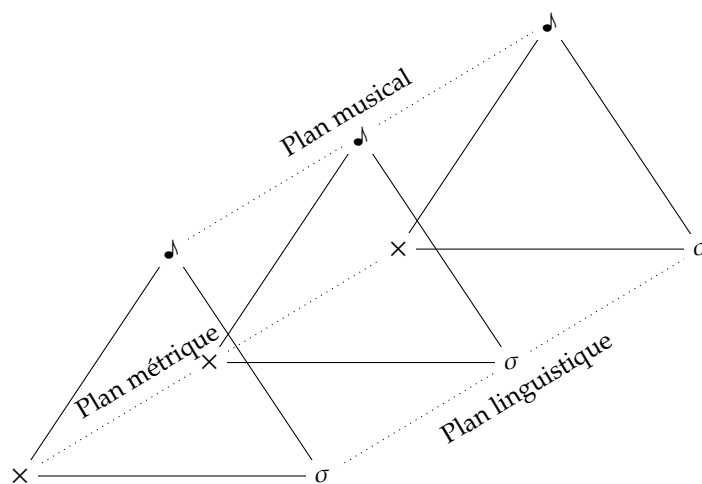


FIGURE 13.1 – Représentation tridimensionnelle du modèle modulaire d'association entre texte, mètre et événements musicaux.

Ce qui importe, en fig. 13.1, c'est que nous concevons que les éléments du plan musical sont associés aux positions métriques et aux syllabes, de manière plus ou moins indépendante, en fonction d'une seconde grammaire d'association. Celle-ci est bien moins sensible à la surjection que ne l'est la grammaire d'association métrique, dans la mesure où les mélismes représentent une situation courante, où une seule syllabe et une seule position métrique correspondent à plusieurs événements musicaux (à plusieurs 'notes'). Nous ne rentrons pas dans les détails de cette grammaire d'association propre à la musication; pour ce travail, ce qui compte est principalement de constater si une syllabe extramétrique dispose, ou non, d'un événement musical associé.⁸ Les deux principes fondamentaux de cette grammaire d'association qui nous sont utiles sont que toute note doit correspondre à une syllabe (13.24) et que toute syllabe doit correspondre à une note (13.25).

(13.24) $\text{♪} \implies \sigma$
 Tout événement musical doit correspondre à une syllabe.

(13.25) $\sigma \implies \text{♪}$
 Toute syllabe doit correspondre à un événement musical.

Dans le cas où une syllabe extramétrique correspond à un événement musical, la situation peut être représentée comme en fig. 13.2 *infra* :

7: PREMAT et CHOUVION 2021; CHOUVION, PREMAT et VERNER 2021; PREMAT, CHOUVION et VERNER 2021.

KIPARKSY (2006). « A modular metrics for folk verse ».

DELL et al. (2009). « Comparing musical textsetting in French and English songs ».

DELL (2015). « Text-to-tune alignment and lineation in traditional French songs ».

8: Suivant l'usage de LERDAHL et JACKENDOFF (1987) parmi d'autres, nous évitons de parler de *note*, qui peut désigner le signe graphique correspondant à un événement musical ou sa hauteur, et qui est de plus anachronique pour nos notations anciennes. Nous ne l'employons que ponctuellement, précisément lorsque la hauteur de la note est un des paramètres étudiés.

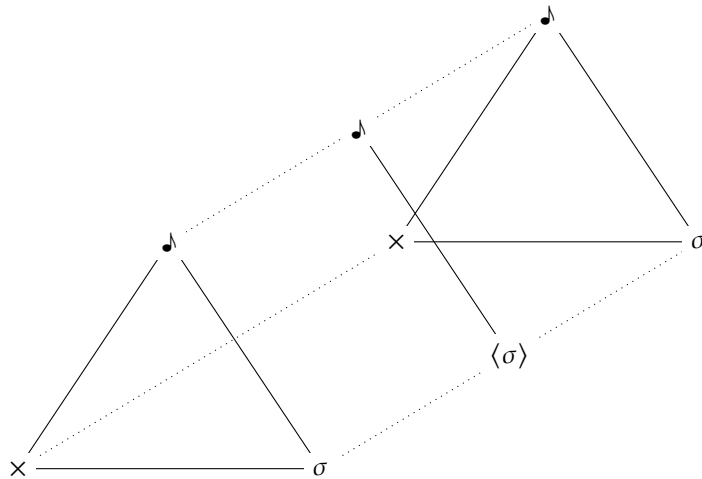


FIGURE 13.2 – Représentation tridimensionnelle du modèle modulaire d'association entre texte, mètre et événements musicaux pour une coupe épique (ici : duplication).

une ligne d'association existe entre un événement musical et la syllabe, mais soit il n'y a pas d'association entre cet événement musical et une position métrique, soit cette association crée un décalage (licite selon la grammaire d'association), l'évènement musical étant associé avec une position métrique associée à une autre syllabe. Ces deux configurations correspondent à ce que nous avons défini respectivement comme une duplication (création d'un nouvel événement musical par copie d'un événement musical adjacent; PREMATA ET CHOUVION 2021, p. 252) et une modification de la constituance neumatique (création, séparation ou modification d'un mélisme, c'est-à-dire d'un groupe d'évènements musicaux correspondant à une seule syllabe; *id.* p. 259).

La représentation musicale de l'extramétricalité est muette, du point de vue analytique, lorsque tous les vers associés à une musique disposent d'une syllabe extramétrique au même droit; c'est le cas, par exemple, lorsque tous les vers correspondant à une phrase musicale sont paroxytoniques. En revanche, le traitement de l'extramétricalité par la musication est apparent lorsque la même phrase musicale est associée à des vers dont la coupe épique ou la terminaison féminine varie selon les vers. Dans ce cas, la notation musicale témoigne d'une modification des événements qui accompagne l'alternance entre terminaisons ou coupes paroxytoniques et oxytoniques. Dans *Aucassin et Nicolette*, par exemple, toutes les laisses sont chantées sur la même musique, mais les laisses peuvent être masculines ou féminines. Pour s'adapter à la présence d'une syllabe de plus dans les laisses féminines, la musique duplique le dernier ou le pénultième événement musical, comme illustré en fig. 13.3 pour les premiers et deuxièmes vers des septièmes, troisièmes et cinquièmes laisses.

VII. Au-ca-ssin s'en est to - nés

III. Au-ca-ssin fu de Biau - cai-re

V. Ni - col' est en pri-son — mi - se

FIGURE 13.3 – « Aucassin et Nicolette », v. 1 et 2 des laisses 7, 3 et 5. Ms. BFN fr 2168, fol. 72r (vii), 70v (iii), 71v (v).

Source du ms. : <https://gallica.bnf.fr/ark:/12148/btv1b9058947p>.

Figures et transcription issues de PREMAY et CHOUVION (2021, p. 257); la seconde portée manque sur le ms. à la strophe v.

Sur cette figure, on observe que les vers de la laisse VII sont masculins ; ils ne comportent que sept syllabes, dont la dernière est associée à la position métrique proéminente du patron métrique (qui ne comporte que sept positions), tandis que les vers des laisses III et V sont féminins, disposant donc d'une syllabe extramétrique en fin de vers. Le premier vers de VII, qui est une laisse masculine, se termine par un neume⁹ complexe (qui comporte deux événements musicaux), un *pes rotundus* \curvearrowright correspondant aux notes [G A],¹⁰ immédiatement précédé par un autre neume complexe, le *climacus* \curvearrowleft correspondant aux notes [A G F].¹¹ Dans les laisses féminines III et V, un neume simple est ajouté, soit avant soit après le *pes rotundus* ; cette note est obtenue par la recopie de l'une des notes du *pes rotundus*. En III, c'est la première note du *pes rotundus* qui est dupliquée : on passe de (13.26) à (13.27). En V, c'est le dernier événement musical du *pes rotundus* qui est dupliqué : on passe de (13.26) à (13.28).

(13.26)	vii	\curvearrowleft \curvearrowright	\curvearrowleft [A G F] \curvearrowright [G A]
(13.27)	iii	\curvearrowleft \curvearrowright \curvearrowright	\curvearrowleft [A G F] \curvearrowright \curvearrowright [G A]
(13.28)	v	\curvearrowleft \curvearrowright \curvearrowright	\curvearrowleft [A G F] \curvearrowright [G A] \curvearrowright

Ce sont des cas que nous avons qualifiés de *duplications régressives* ou *progressives* (PREMAY et CHOUVION 2021, p. 252-253), selon que la duplication recopie le premier événement musical du dernier neume complexe ou le dernier événement musical de la phrase. Cette situation correspond donc au schéma en fig. 13.2 *supra* : une note a été ajoutée pour la syllabe extramétrique, et cette note ne correspond à aucune position métrique. Lorsque la duplication est régressive (iii), cela décale l'association entre notes et positions métriques : ce n'est plus le *pes rotundus* \curvearrowright [G A] qui correspond à la position proéminente du gabarit métrique, mais c'est le neume simple \curvearrowright issu de la duplication.

Une duplication s'observe également entre les deuxièmes vers des laisses VIII et V. Cette fois-ci, il n'y a pas de neume complexe, la phrase masculine

9: Le mot *neume* est utilisé, dans ce document, pour désigner le signe graphique représentant un ou plusieurs événements musicaux (ce qu'on appelle une *note* en solfège moderne). Un neume peut induire une constituance (dans le cas des neumes complexes), des informations de hauteur de note, de rythme, de proéminence, etc.

10: Comme c'est l'usage en musicologie médiévale, nous utilisons les noms alphabétiques des hauteurs de notes, réservant les noms habituels du solfège français Ut (= Do) Ré Mi Fa Sol à la solmisation, pratique qui permet de placer les intervalles dans le mode (la note La n'existe pas au Moyen Âge).

11: La constituance neumatique est rendue par des ligatures dans notre transcription et par des crochets en notation alphabétique. Dans la notation médiévale, cette constituance est sans ambiguïté : les événements musicaux formant un neume complexe doivent être chantés sur la même syllabe.

(v) étant entièrement syllabique (un neume simple par syllabe). La duplication recopie, ici, le dernier neume simple : au lieu de se terminer sur B G (13.29), la phrase musicale correspondant au vers féminin se termine sur B G G (13.30), et l'on ne peut évidemment savoir si la duplication est régressive ou progressive.

(13.29) VII ■■ ■^B ■^G
 (13.30) III ■■■ ■^B ■^G ■^G

Cette situation étant assez courante dans les sources médiévales,¹² elle semble tout à fait licite du point de vue de la grammaire d'association ; c'est pourquoi nous n'avons pas postulé de règle régissant l'association entre événements musicaux et positions métriques en (13.24 - 13.25) *supra*.¹³ Elle témoigne du fait que l'alternance entre terminaisons masculines et féminines est prise en charge sans difficulté par la grammaire d'association de la musication. Dans aucun cas nous n'avons rencontré de vers féminins notés directement sous la musique et dont la musique ne témoigne pas de la mise à disposition d'un événement musical pour la posttonique.

13.1.4 Échec heuristique du PAM

Le projet originel de cette thèse reposait sur l'usage du PAM, dont le but était de découvrir des irrégularités métriques imputables à des traitements imprédictibles du schwa final dans un large corpus de textes médiévaux. Loin de permettre ce type de découvertes, le PAM a principalement conforté le dogme de l'élision décrit *infra*. Dans une écrasante majorité des textes de toutes époques, le PAM n'a détecté que quelques pourcentages de vers anisosyllabiques, et ces cas sont principalement dus soit à un mauvais paramétrage linguistique du PAM (qui ne peut prendre en compte toutes les variations graphiques sans l'intervention de l'utilisateur pour compléter ses dictionnaires d'exceptions) soit à des vers dont l'irrégularité est orthogonale au traitement d'un schwa final. Si, dans les textes anglo-normands, le PAM découvrait des irrégularités plus prononcées, celles-ci ne tenaient jamais significativement au traitement d'un schwa final. Nous avons été jusqu'à produire un ensemble de mesures exploratoires pour tenter de trouver quels facteurs liés aux voyelles élidables pourraient entrer en jeu, mais les résultats se sont trouvés réduits aux cas d'élision non transcrite dans *Roland* et *Alexis* (*ço est pour cest, jo ai pour jai, li emperere pour l emperere*, etc.)¹⁴ et à quelques cas où c'est le maintien de la voyelle qui est attendu par le mètre alors que la graphie est élidée (p.ex. *Carles l emperere pour Carles li emperere, Roland*, v. 2753).

Si la documentation de ces cas d'élisions non transcrites et de dénis d'élisions pourtant transcrites ont une certaine valeur heuristique, elles demeurent confinées à des textes anglo-normands, pour lesquels l'anisosyllabisme n'est pas aussi exceptionnel que sur le continent. En revanche, en ne trouvant pas d'irrégularité significative sur le continent, le PAM a fait l'inverse de ce pour quoi il était conçu : il a démontré la régularité à toute épreuve de l'élision métrique, telle que détaillée *infra*. De fait, nous ne disposons pas des résultats définitifs du déploiement du programme sur l'ensemble du corpus de la BFM ; la phase exploratoire, qui a consisté

12: Cf. d'autres exemples tiré du *Gieus de Robin et Marion* (ADAM DE LA HALLE 2002 [1998]), v. 1-4 et 97-98, analysés dans PREMAT et CHOUVION (2021, p. 258).

13: Ceci est probablement une simplification, mais elle est suffisante pour notre étude, qui demeure très en surface sur ce point.

14: PREMAT et POGGIO ([à paraître]) :

- *Ço est pour cest* : *Roland* v. 490, 786, 866, 1433, 1527, 2236, 2613, 3208, 3228, 3224, 3258, 3309, 3778 ; *Alexis* v. 49, 178, 359, 363, 436, 441.
- *Jo + « avoir »* : *Roland* v. 2404, 2408, 2768, 2903.
- *Li emperere(s) pour l emperere(s)* (CSsg) : *Roland* v 180, 274, 383, 740, 783, 1832, 2352, 3119, 3841, 3970.
- et d'autres : *que anme pour qu anme* (*Roland* v. 1846), *que en pour qu en* (*Roland* v. 197, 837, 2437, 2687), *se il pour s il* (*Roland*, v. 2134), *se a(d) pour s a(d)* (*Alexis*, v. 151, 285), etc. *Alexis* = **AlexisRAM** : ANON. (1050*). *Vie de saint Alexis*. Compo. ca 1050, ms. ca 1120, dial. norm. (RAINSFORD et MARCHELLO-NIZIA 2018).

à exécuter le programme sur 54 textes, s'est révélée suffisamment décevante pour que nous n'allions pas plus loin dans cette voie. De fait, c'est ce constat, contraire à ce que nous espérions trouver, qui fonde les données sur l'élision métrique de ce chapitre : celle-ci est régulière dans une écrasante majorité des cas. Des graphiques permettant de documenter la démarche exploratoire sont rejetés en annexes pour *Alexis* et *Roland* (fig. E.24 - E.26) ; leur lecture n'informe en rien la lecture de ce chapitre, mais documente l'absence d'effet patent de la présence de schwas finaux sur l'irrégularité des vers.

13.2 L'élision métrique

Cette section est dédiée à l'élision métrique, c'est-à-dire à l'élision telle qu'on peut la reconstituer pour faire correspondre le vers à son patron métrique. Elle s'ouvre par le constat que la *doxa* des manuels sur l'élision est fiable dans les grandes lignes, et que tout schwa final atone placé au sein d'un constituant métrique s'élide (13.2.1). Elle envisage ensuite le traitement musical de l'élision métrique, qui suit la plupart du temps l'élision que l'on peut inférer de la seule structure métrique (sec. 13.2.1.1), avant d'interroger le conflit possible entre coupe épique et élision (sec. 13.2.2), puis de formaliser la question du domaine de l'élision métrique (sec. 13.2.3).

13.2.1 L'élision métrique médiévale

Pour obtenir des scores élevés de régularité vis-à-vis de leur patron métrique et *via* la grammaire d'association ébauchée en (13.20 - 13.23), les textes répondent à la règle d'élision en (13.2.2), qui est celle du français classique. La règle, couramment énoncée par les manuels de métrique, est valide dans nos données — à l'exception de ses conséquences sur la coupe épique. Avant d'aborder le point de la coupe épique, observons la règle telle qu'énoncée par Mazaleyrat, ce qui nous permet de démontrer que, dans les grandes lignes, la règle est bien énoncée par les grammaires scolaires.

Problème de l'élision

Il est résolu dès l'origine : l'*e* terminal d'un mot suivi d'une initiale vocalique ou d'un *h* non aspiré n'a pas plus d'existence rythmique qu'il n'en a de phonétique :

Vostr(e) olifant ne deignastes suner. *Chanson de Roland*

La règle ne changera pas.

(MAZALEYRAT 1990 [1965], p. 59)

MAZALEYRAT (1990 [1965]). *Éléments de métrique française*.

La règle donnée par Mazaleyrat suppose donc que le schwa final de *vostre* est absent à la fois au niveau métrique et au niveau de la prononciation. En éditant le vers pour remplacer le <e> élidé par une apostrophe, on obtient la conformité au patron métrique donnée en (13.31).

(13.31) × × × '× × × × × × '×
 | | | | | | | | | |
Vostr' o- li- fan ne dei- gnas- tes sun- ner
Votre olifan ne deignastes sunner
 « Votre olifan, vous n'avez pas daigné le sonner »
 (*Roland*, v. 1171, trad. SHORT 1990)

Ce type d'élision ne concerne pas que les mots grammaticaux (*vostre* en 13.31); tout schwa final de mot polysyllabique est sujet à l'élision devant initiale vocalique. Ainsi, au v. 3 de *Roland* (13.32), l'élision atteint l'adverbe *tresque* devant *en* (→ *tresqu'en*) et le substantif *tere* devant *altaigne*; dans le second cas, l'élision n'est pas transcrite. Cette règle est générale pour les polysyllabes, sans quoi les résultats du PAM n'atteindraient pas de tels scores de conformité au patron métrique.

(13.32) × × × '× × × × × × '×
 | | | | | | | | | |
Tres- qu'en la mer con- quis- la ter' al- tai- <gne>
Tresqu'en la mer cunquist la tere altaigne
 « Jusqu'à la mer il a conquis les terres hautes »
 (*Roland*, v. 3, trad. SHORT 2007)

Pour les monosyllabes, en revanche, la règle est moins régulière. Puisque seules les voyelles finales atones s'élident, les monosyllabes sujets à l'élision ne peuvent être que des monosyllabes fonctionnels. Or nous avons vu au chapitre précédent que l'élision graphique de certains d'entre eux n'était pas toujours systématique, avec notamment « que » (élément *que*, mélangé à des pronoms potentiellement toniques, ou locutions en *base + -que*), *je* et *li* sg. La citation de Mazaleytrat continue, en-deçà du passage cité *supra*, en mentionnant précisément un cas de résistance de *que* à l'élision :

La règle ne changera pas. Et les survivances d'*e* non élidés qui subsisteront un temps dans quelques cas d'emplois de pronoms, articles, adverbes ou conjonctions :

Qu'en dites vous? quE il vos semble? RUTEBGEUF
 (élision) (e non élidé)

ne restreignent de cette règle ni l'application, qui est générale, ni la valeur, qui est définitive : elle a toujours été et elle reste une des constantes du système prosodique français.
 (MAZALEYTRAT 1990 [1965], p. 59)¹⁵

MAZALEYTRAT (1990 [1965]). *Éléments de métrique française*.

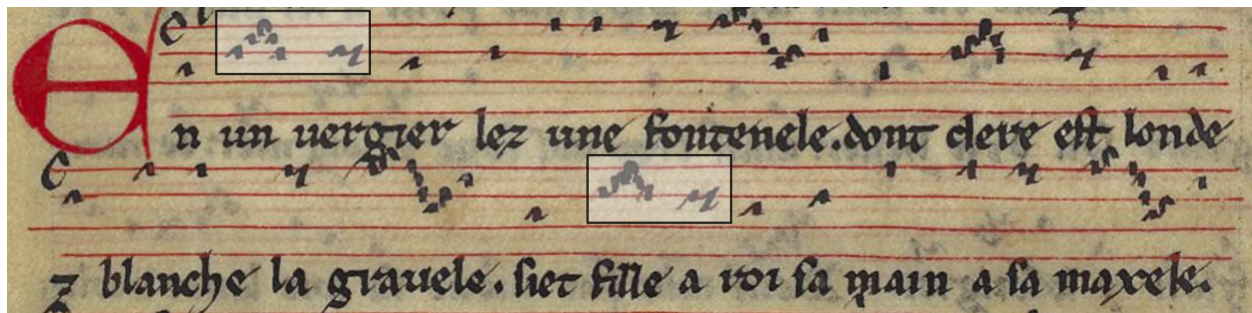
15: L'emphase sur « quE » est de l'auteur.

Ces cas de déni d'élision échappent au PAM, du moins si la graphie les transcrit : le PAM n'élide pas les monosyllabes. Mais, de fait, ces cas n'échappent pas à notre analyse : ils sont alors repérables à la graphie, et donc relevés par Comparalem. Quant aux cas où un monosyllabe ne transcrirait pas l'élision alors que celle-ci est nécessaire à la conformation au patron métrique, ils sont détectés par le PAM sous la forme d'une hypermétrie; *Roland* atteste ainsi d'occurrences *le* pour *l*, *que* pour *qu*, *se* pour *s*, *ço* pour *c*, *jo* pour *j* (cf. sec. 9.1.1.1 et MARCHELLO-NIZIA 2015), *ki* pour *k* et *li* (sg) pour *l*. Nous avons jusqu'à présent conservé une

certaine prudence quant à la représentativité phonologique de ces élisions métriques et non graphiques, considérant que la phonologie du vers n'était pas nécessairement un témoin fiable de la phonologie de la langue. Avant de développer ce point, nous reprenons la question de l'interaction entre élision métrique et coupe épique.

13.2.1.1 Élision métrique et musication

Si l'on peut rencontrer ponctuellement des cas où la musique semble accueillir une finale atone qui devrait être élidée, ces cas demeurent exceptionnels. Dans une écrasante majorité des cas, la musique respecte l'élision des voyelles à l'intérieur du vers et/ou de l'hémistiche telle que décrite *supra* (cf. sec. suivante pour la question du vers ou de l'hémistiche). C'est par exemple le cas dans la chanson *En un vergier*, où la même phrase musicale se répète à l'identique sur les vers 1 et 3, le v. 3 contenant une syllabe régulièrement élidée. En fig. 13.4, les neumes correspondant au *locus* de l'élision au v. 3 et à son équivalent au v. 1 sont encadrés.



1. En un _____ ver - gier
3. Siet fill' _____ a _____ roi,

FIGURE 13.4 – « En un vergier », v. 1 et 3. Ms. U, fol. 65v.
Source du ms. : <https://gallica.bnf.fr/ark:/12148/btv1b60009580/f138.item>.
Figure et transcription issues de PREMAT et CHOUVION (2021, p. 253).

Sur la fig. 13.4, on observe que la phrase musicale n'est pas modifiée pour s'adapter à la présence de la voyelle élidée de *fil*le *a* : la musique est la même pour *En un vergier* et pour *siet fille a roi*. L'élision métrique du schwa final de *fil*le est donc respectée par la musique également, qui n'insère pas un nouveau neume ni ne détache l'un des constituants des neumes complexes environnants — ce qu'elle ferait pour accueillir une terminaison féminine ou une coupe épique. La graphie musicale indique donc que le mélisme [A B C B A] correspond à la première syllabe de *fil*le et que la *clivis* ♯ [A G] subséquente correspond à la syllabe *a*; aucun neume (et donc aucun évènement musical) n'est disponible pour la posttonique élidée. C'est cette situation, l'absence de mise à disposition d'un évènement musical pour les voyelles élidées, qui est la norme dans les répertoires que nous avons étudiés.

13.2.2 Élision métrique et coupe épique

Nous adoptons provisoirement les définitions suivantes de l'élision métrique, la première pour les polysyllabes et la seconde pour les monosyllabes :

- (13.33) ÉLID($\sigma\sigma$)
Toute voyelle finale atone de polysyllabe immédiatement suivie d'une initiale vocalique s'élide.
- (13.34) ÉLID(σ)
Toute voyelle finale atone de monosyllabe immédiatement suivie d'une initiale vocalique peut s'élider.

Ces définitions sont celles qui fondent l'algorithme du PAM (cf. sec. 3.3). La première est celle que le PAM calcule lui-même à partir des propriétés d'adjacence des mots, des étiquettes de son annotation et de la présence ou absence d'une consonne entre les deux mots. La seconde n'est pas traitée par le PAM lui-même, il la délègue à la graphie, de sorte que la donnée est portée par l'entrée du programme et n'a pas besoin d'être travaillée par le programme. Ces définitions font la prédiction suivante : toute voyelle finale atone en fin d'hémistiche devrait être élidée si l'hémistiche suivant commence par une voyelle. Elles prédisent donc l'absence théorique de coupes épiques devant initiale vocalique, c'est-à-dire de syllabes extramétriques en fin d'hémistiche ailleurs que devant consonne.

13.2.2.1 Coupe épique devant initiale consonantique

De fait, bien des manuels traitent la coupe épique comme un cas de suppression pure et simple de la voyelle posttonique. Ainsi, Mazaleyrat toujours, classe le v. 16 de *Roland*, représenté en (13.18) *supra* parmi les cas « d'apocope », c'est-à-dire de « disparition en fin de mot d'un *e* non élidable » (MAZALEYRAT 1990 [1965], p. 60). D'une part cette définition de ce que nous appelons coupe épique fait une prédiction sur la prononciation du vers que nous pensons fautive, d'autre part elle restreint le domaine de la coupe épique aux contextes où l'élision n'est pas possible. Aussi, au lieu de (13.35), reproduction de (13.18), l'application de « l'apocope » produirait (13.36).

- (13.35) × × × '× × × × × × '×
| | | | | | | | | |
Li em- pe- re- <res> Car- les de Fran- ce dul- <ce>
« Charles, l'empereur de France la douce »
(*Roland*, v. 16, trad. SHORT 1990)

- (13.36) × × × '× × × × × × '×
| | | | | | | | | |
Li em- pe- rer' Car- les de Fran- ce dul- <ce>

Non seulement le syntagme *li emperer'* est probablement agrammatical en ancien français, mais il découle d'une théorie du vers français qui écrase forme de surface et forme sous-jacente sur le modèle de la prononciation

moderne des vers, ce qui est à la fois anachronique et simpliste. Il semble préférable, avec CORNULIER (2012), d'admettre qu'un décasyllabe de *Roland* peut être prononcé en 10, 11 ou 12 syllabes, mais dont seules 10 ont un rôle métrique.

Les coupes épiques, tout comme les terminaisons féminines de vers, sont également attestées dans le répertoire musical. Dans notre corpus de 157 chansons de trouvères¹⁶, les occurrences sont peu nombreuses, mais les témoignages que nous avons repérés sont édifiants.

13.2.2.2 Élision de coupe épique devant initiale vocalique

Si l'on suit la logique de Mazaleyrat et de bien d'autres, il ne peut donc exister de coupe épique devant initiale vocalique, puisque l'élision dispense de recourir à ce qu'il qualifie d'apocope. De ce fait, le v. 10 de *Roland* ne correspondrait pas à (13.37), que dérive notre grammaire d'association, mais à (13.38), où la dernière syllabe de *Marsilie* (paroxyton trysyllabique) serait élidée.

(13.37) × × × '× × × × × × '×
 | | | | | | | | | |
 Li reis Mar- si- <lie> es- teit en Sar- ra- gu- <ce>
 « Le roi Marsile se tient à Saragosse »
 (*Roland*, v. 10, trad. SHORT 1990)

(13.38) × × × '× × × × × × '×
 | | | | | | | | | |
 Li reis Mar- sil' es- teit en Sar- ra- gu- <ce>

Cette prédiction vient en contradiction avec ce que prédit la grammaire d'association en (13.20 - 13.23), qui tendrait plutôt à considérer que la coupe épique neutralise la possibilité de l'élision en plaçant chaque hémistiche dans un domaine computationnel étanche. Aussi, si la coupe épique est une fin de domaine, alors *Marsilie* n'est pas suivi par une initiale vocalique, et l'élision est hors de propos.

Modification de la constituance neumatique. Lorsque la source est unique, elle indique parfois une stratégie de redistribution des événements musicaux qui vise à rendre un événement disponible pour la syllabe extramétrique. C'est le cas en fig. 13.5, qui représente les vers 1 et 3 de la chanson *Bele Doette*. Le v. 1 dispose d'une coupe épique devant initiale vocalique (*Bele Doet<te> / as fenestres se siet*), tandis que le v. 3 dispose d'une coupe oxytone (*de son amy / Doon li resovient*).¹⁷

CORNULIER (2012). « Si le mètre m'était compté... Sur la notion fallacieuse de mesure du vers ».

16: Dont nous n'avons étudié que la musique notée directement sous la partition, c'est-à-dire le plus souvent la première strophe; cf. sec. 2.3.1.5.

17: Le mot *amy* est surscrit dans le ms.



FIGURE 13.5 – « Bele Doette », v. 1 et 3. Ms. U, fol. 66r.
 Source du ms. : <https://gallica.bnf.fr/ark:/12148/btv1b60009580/f139.item>.
 Figure issue de PREMAT et CHOUVION (2021, p. 260).

La musique du premier vers contient une séquence C C D, dont le premier C correspond à la tonique de *Doette* et le second à la posttonique extramétrique, et dont le D correspond à la première syllabe du second hémistiche *as*. Le second vers contient une séquence C [C D], dont les deux dernières unités forment un neume complexe, le *pes rotundus* ♫, correspondant à la première syllabe du second hémistiche. Ces séquences sont encadrées en fig. 13.5, et illustrent l'un des mécanismes par lequel la musique s'adapte à la présence d'une syllabe extramétrique. Ainsi, l'opposition entre, d'une part trois *virgae* ♣ ♣ ♣ correspondant à *(Do)et<te> as*, et d'autre part une *virga* et un *pes rotundus* ♣ ♫ correspondant à *(a)-my Do(on)* fait clairement signe vers le fait que la coupe épique devant initiale vocalique était chantée comme telle. Le phénomène ici à l'œuvre n'est pas la duplication comme illustré schématiquement en fig. 13.2, mais une modification de la constituance neumatique : la musique s'adapte à l'alternance entre coupe masculine (*amy*) et épique (*Doette*) en jouant sur la constituance d'un neume. Il convient donc de poser, pour ce vers, la correspondance entre texte et mètre telle qu'en (13.39), où la posttonique de *Doette* n'est pas élidée mais est extramétrique, et non (13.40), où la posttonique serait élidée.¹⁸

(13.39) × × × '× × × × × × '×
 | | | | | | | | | |
 Be- le Do- et- <te> as fe- nes- tres se siet
 « La Bele Doette s'assied devant les fenêtres »
 (« Bele Doette », v. 1, trad. pers.)

(13.40) × × × '× × × × × × '×
 | | | | | | | | | |
 Be- le Do- et' as fe- nes- tres se siet

18: Pour un exemple analogue dans la chanson de trouvère « Amors qui m'a en baillie », cette fois-ci devant initiale consonnantique, cf. PREMAT et CHOUVION (2021, p. 260-262).

Duplication. L'autre stratégie d'adaptation de la musique à la présence d'une syllabe extramétrique évoquée en sec. 13.1.3.3 peut aussi être présente pour des coupes épiques devant initiale vocalique. C'est le cas dans la chanson « En un vergier », dont les trois premiers vers correspondent à la même phrase musicale, mais dont le second comporte une coupe épique devant initiale vocalique : *Dont clere est l'on(de) et blanche la gravele*. Comme dans l'exemple précédent, la musique s'adapte à cette coupe épique, mais elle ne le fait plus en modifiant la constituance neumatique. Ici, la mélodie répète l'un des neumes simples. Les trois sections de la notation correspondant à la coupe de ces trois vers sont encadrées en fig. 13.6.

FIGURE 13.6 – « En un vergier », v. 1 et 3. Ms. U, fol. 65v.

Source du ms. : <https://gallica.bnf.fr/ark:/12148/btv1b60009580/f138.item>.

Figure et transcription issues de PREMAY et CHOUVION (2021, p. 253).

Ici, c'est par l'alternance entre *clivis - virga* ♪ ♫ et *clivis - virga - virga* ♪ ♫ ♫ que le ms. signale que, bien que placée devant initiale vocalique, la voyelle finale de *onde* est bien chantée. Il y a donc eu duplication, du premier neume du second hémistiche ou du dernier constituant de la *clivis* correspondant à la tonique de l'hémistiche.

Ambiguïté entre les sources. Dans d'autres cas, néanmoins, il peut y avoir ambiguïté. C'est par exemple le cas au v. 7 de la chanson « La douce voiz du rosignol sauvage », chanson qui nous est transmise par 5 ms. (O, M, P, X et a). Le vers est *bien doi avoir grant joie en mon corage*, et peut s'interpréter comme un décasyllabe à césure 4/6 (la césure est localisée après la quatrième position métrique, comme en 13.10) soit comme un décasyllabe à césure 6/4,¹⁹ auquel cas *joie* est le dernier mot du premier hémistiche. Les ms. PXa disposent d'un neume de plus que les ms. OM, et ce neume supplémentaire correspond au *-e* de *joie*.

Si l'on considère que ce décasyllabe est bien à césure 6/4, ce que transcrivent les ms. PXa, c'est une duplication — comme en fig. 13.6 *supra*, la musique s'adapte pour accueillir la syllabe extramétrique (13.41). Toujours dans cette hypothèse de césure 6/4, les ms. OM refusent cette

19: Cette variante du patron métrique 10s en (13.10) est acceptée comme bien formée par une partie des textes médiévaux (LOTE 1949-1996 [1991], t. 1, 2, chap. 1, §118).

FIGURE 13.7 – « La douce voix du rossignol sauvage », v. 7. Ms. O, fol. 74v; M, fol. 54v; P, fol. 33v; X, fol. 71v; a, fol. 13r. Sources des ms. en annexes. Figure issue de PREMAY et CHOUVION (2021, p. 260). Les images des ms. O, P et X ont été éditées pour rajouter la clef d'Ut manquante, sous la forme d'un 'C'.

coupe épique et pratiquent l'élision du *-e* de *joie*, ne laissant aucun événement musical disponible pour la syllabe finale de *joie* (13.42). Au moins trois autres hypothèses peuvent être considérées. D'une part, on peut envisager le fait que les scribes de PXa autorisent la césure 6/4 (ce qui entraîne la reconnaissance de l'extramétricalité et la duplication; 13.41) tandis que les scribes de OM analysent ce vers comme étant un décasyllabe à césure 4/6, ce qui entraîne automatiquement l'élision de *joie*, le mot n'étant plus en fin d'hémistiche (13.43). D'autre part, on peut envisager que dans tous les ms. le vers est un décasyllabe à césure 4/6, et que c'est par exception que les scribes de PXa décident de ne pas éluder le *-e* de *joie* (13.44). Comme mentionné *supra*, de telles exceptions sont rares si ce n'est marginales; elles procèdent probablement du fait que la musique vocale dispose d'une grammaire d'association différente de celle de la métrique, plus libre sur certains points et plus contrainte sur d'autres.²⁰ Enfin, le mot *joie* comportant un hiatus $V\emptyset$, il est peut-être concerné par la réduction des schwas finaux en hiatus (É. BOURCIEZ 1967 [1910], p. 644; ZINK 2013 [1986], p. 47; ENGLEBERT 2015, p. 93; PIERRET 1994,

20: DELL (2015) postule l'absence totale de patron métrique dans la chanson traditionnelle française. Nous n'acceptons pas cette conclusion pour notre corpus, dont les textes sont très majoritairement conformes aux patrons métriques.

p. 218-219). Dans ce cas, les scribes de PXa n'appliquent pas la réduction du hiatus (13.44), par abstention volontaire ou parce que leur grammaire n'est pas encore atteinte par cette évolution, tandis que les scribes de OM appliquent la réduction de l'hiatus final et, peu importe où tombe la coupe, ils ne peuvent maintenir le *-e* de *joie*.

- (13.41) PXa *bien doi avoir grant joi⟨e⟩ / en mon corage*
 (13.42) OM *bien doi avoir grant joi' / en mon corage*
 (13.43) OM *bien doi avoir / grant joi' en mon corage*
 (13.44) PXa *bien doi avoir / grant joie en mon corage*

Il est difficile de privilégier l'une ou l'autre de ces hypothèses. C'est sur ce type d'exemple que s'appuie une partie de la bibliographie philologique pour postuler que les coupes épiques sont absentes du répertoire des trouvères et que l'apparition d'une note pour une finale qui devrait être en élision métrique démontre simplement la liberté de l'association entre texte et musique vis-à-vis du mètre.²¹ Dans notre corpus, ces configurations sont néanmoins plus rares que celles où la musique indique le respect de la coupe épique ; l'exemple de *joie* en fig. 13.7 est le seul que nous ayons repéré dans nos 157 chansons de trouvères.

21: Pour un aperçu du débat, cf. SANGIOVANNI (2020).

13.2.2.3 En somme

Devant initiale consonantique, la syllabe atone finale en coupe épique ne peut être sujette à l'élision. Plutôt que de supposer qu'elle est supprimée par apocope, nous adoptons la position selon laquelle cette syllabe est extramétrique, c'est-à-dire réalisée mais non associée à une position métrique. La musique témoigne également de ce fait, dans la mesure où elle rend disponible un évènement musical qui permet la réalisation de la syllabe extramétrique. Devant initiale vocalique, l'élision est possible, mais n'est guère attestée en métrique seule (nos sources métriques n'ont pas les moyens d'en attester), tandis que la musication témoigne majoritairement du même accommodement de la mélodie à la présence d'une syllabe extramétrique que devant initiale consonantique. Nous conservons donc l'analyse dans laquelle une finale extramétrique ne s'élide pas, même si l'hémistiche suivant commence par une voyelle.

13.2.3 Domaine où l'élision métrique a lieu

Sur la base des remarques de la section précédente, nous pouvons délimiter le domaine métrique dans lequel s'applique l'élision : il correspond au vers ou, lorsqu'une coupe épique est admise par le texte, à l'hémistiche. Ces domaines, purement métriques, ne sont pas réductibles à des domaines linguistiques. S'il existe des grammaires d'association qui reposent sur l'idée d'une congruence entre constituants syntaxiques et métriques (p.ex. DELL et BENINI 2021), l'analyse habituelle de ces domaines se fait, dans le cadre de la phonologie prosodique classique (SELKIRK 1980 [1972], 1981 ; NESPOR et VOGEL 2007 [1986]), en implémentant dans la grammaire d'association des conditions d'association entre constituants prosodiques et constituants métriques. La question de l'identité représentationnelle à donner au domaine n'est pas central pour nous ; nous admettons simplement l'idée que la métrique et la musication procèdent

SELKIRK (1980 [1972]). *The Phrase Phonology of English and French*.
 — (1981). « On prosodic structure and its relation to syntactic structure ».
 NESPOR et VOGEL (2007 [1986]). *Prosodic Phonology*.

à des alignements entre les domaines phonologiques et les constituants métriques et musicaux.

Ce dont témoignent les résultats du PAM et la musication, c'est la généralisation de l'élision métrique telle qu'énoncée en (13.33 - 13.34) *supra*, à ceci près que l'élision est bornée par les frontières du vers (elle ne frappe pas les terminaisons féminines de vers) et, dans le cas des coupes épiques, par les frontières de l'hémistiche. Le domaine de l'élision peut donc être résumé à la généralisation en (13.45). Dans la mesure où celle-ci repose sur la désignation de constituants, on peut également obtenir une formulation, certes plus complexe, mais qui, en conformité avec les patrons dépourvus de frontières proposés en (13.9 - 13.11), ne fait référence qu'aux propriétés des positions, tel qu'en (13.45).

(13.45) ÉLID.DOM.^{diac.}

L'élision métrique a pour domaine un constituant métrique.

(13.46) ÉLID.DOM.

$$\{\emptyset, V\} \rightarrow \emptyset \neg / \left[\begin{array}{c} ' \times \\ \vdots \\ ' \sigma \quad _ \end{array} \right]$$

Une syllabe phonologiquement élidable subit l'élision métrique lorsqu'elle n'est pas (¬) la posttonique d'un mot dont la tonique est associée avec une position proéminente.²²

L'énoncé formel en (13.46) présente exactement le même contexte que celle de la règle d'association entre syllabes et positions en (13.22) *supra*, contexte que nous reproduisons en (13.47). Cette règle autorise, en négatif, l'extramétricalité des posttoniques dont la tonique est associée à une position proéminente, puisqu'elle n'interdit l'extramétricalité que pour les autres syllabes atones (la règle $'\sigma \Rightarrow \{ \times, ' \times \}$ en 13.23 *supra* gère les syllabes accentuables).

$$(13.47) \sigma \Rightarrow \times \quad \neg / \left[\begin{array}{c} ' \times \\ \vdots \\ ' \sigma \quad _ \end{array} \right]$$

Cette communauté de contexte nous assure que les seules syllabes finales atones à échapper à l'élision sont celles qui sont extramétriques, ce qui génère les configurations attestées dans nos données. Du fait de la restriction contextuelle de (13.45), les finales atones de fin d'hémistiche (lorsqu'il y a coupe épique) et de fin de vers sont protégées de l'élision, mais ne sont pas métriquement inconvenantes non plus puisqu'elles ne sont pas concernées par la règle d'association entre syllabes et positions (13.47) et peuvent donc demeurer flottantes.

13.3 Naturalité du domaine de l'élision métrique, parophonologie et conflit

Considérer que c'est le domaine métrique qui, en vers, définit les bornes entre lesquelles s'applique l'élision métrique nous paraît naturel à au moins deux titres. Par naturel, nous entendons qu'il peut être assez simplement dérivé d'autres propriétés et n'est pas entièrement arbitraire.

22: Phonologiquement élidable : finale atone d'un polysyllabe ou d'un monosyllabe, ayant été déclarée élidable par le test pré-historique COND CV (14.1) et se trouvant devant un mot à initiale vocalique.

D'une part, tel que formulé en (13.46), le domaine dérive directement du gabarit métrique : ce sont les positions proéminentes '× qui définissent les frontières du domaine. Autrement dit, la métrique ne peut définir un domaine plus petit que l'hémistiche (ou le vers, dans le cas des vers simples) dans lequel s'appliquerait la règle de l'élision, dans la mesure où elle ne dispose d'aucun domaine plus petit. Si l'on admet, sur le fondement des généralisations graphiques du chapitre précédent, que le domaine de l'élision ne comprend qu'exceptionnellement plusieurs mots lexicaux, il faut alors constater que la métrique *force* le domaine de l'élision à s'élargir pour correspondre au plus petit domaine qu'elle connaît. Ce tour de force de la métrique tient à sa dimension paraphonologique, développée en sec. 13.3.2.

Par ailleurs, et c'est le premier point que nous traitons (sec. 13.3.1), si nous avançons que le domaine de l'élision métrique est naturel, c'est aussi qu'il correspond assez bien à la prosodie de la fin du Moyen Âge et de la Renaissance, telle qu'attestée par les grammairiens et métriciens du xvi^e s.

13.3.1 Congruence entre l'élision métrique et celle du xvi^e s.

Les témoignages des grammairiens du xvi^e s. renseignent non seulement sur les lieux où s'applique l'élision, mais également sur les lieux où elle ne s'applique pas, définissant ainsi son domaine par l'intérieur et par les frontières. Ainsi, dans son manuel d'apprentissage du français à l'intention d'un public anglophone, Palsgrave décrit que le schwa final disparaît sous l'effet de la fusion (« as though they made both b[y] one word[...] ») entre les deux mots subséquents, sauf en fin de phrase :

But²³ if a frenche worde ende in *e*, nat havynge his accent upon the same *e* nor beyng a femine participle, the next worde folowyng begynnynge also with a vowel, eyther alone or in a diphtong, or with *h* at havynge his aspiration, than shal the *e* of the worde so goyng before lese his sounde, and the consonant comyng before *e* be joyned in sound to the vowel folowyng, as though they made both bu one worde.

Except (as I have sayd) the comyng before a poynt cause the contrary, as this sentence *Il a trovue une aultre dame belle et gracieuse* shall in redyng and spekyng be sounded : *Il a trouue un autre dame bellegracieuse*; and so of all other.

(PALSgrave 1530, p. 41)

23: Au paragraphe précédent, Palsgrave explique que <e> final se maintient lorsqu'il est accentué.

PALSgrave (1530). *L'Esclaircissement de la langue françoise*.

Palsgrave a donc l'intuition du *sandhi* externe qui relie les deux mots et génère l'élision. La règle qu'il donne est très proche de celle que nous avons donnée pour l'élision métrique : au sein d'un domaine pouvant contenir plusieurs mots lexicaux, les finales atones disparaissent devant initiale vocalique (*une autre* → *un autre*; *belle et gracieuse* → *bellegracieuse*). Ronsard, ne parlant que des règles du vers, ajoute que des voyelles autres que schwa peuvent être sujettes à l'élision, ce qui correspond aux données synthétisées au chapitre précédent :

DE LA VOYELLE, E

Toutes fois & quantes que la voyelle, e, est rencontrée d'un autre voyelle ou diftongue, elle est toujours mangée, se perdant en la voyelle qui la fuyt, sans faire syllabe par foy, ie dy rencontrée d'une voyelle ou d'une diftongue pure, autrement elle ne se peut manger qu'ad l'i & v voyelles se tournent en consonnes, cōme, *le viue*, Exemple de e, qui se mange, *cruelle & fiere, & dure, & facheuse amertume. Belle maistresse inexorable & fiere*. Davantage i, & a, voyelles se peuvent elider & manger. Exemple d'a, *L'artillerie, l'amour*, pour *la artillerie, la amour*, Exemple de la voyelle i, *n'a ceux cy, n'a ceux là*, pour dire *ny a ceux cy ny à ceux là [...]*

(RONSARD 1585, p. 36-37)²⁴

Si Ronsard ne commente ni le statut monosyllabique ni le statut grammatical des mots dont il élide <a> et <i>, sa conclusion est conforme à nos données sur ce point. En revanche, comme Palsgrave et comme nos données sur l'élision métrique, Ronsard indique l'élision systématique de la finale des paroxytons sans consonne finale devant initiale vocalique, que ce soit dans une structure *Nom - Conj* (*cruelle & fiere*), *ADJ - NOM* (*faceuse amertume*) ou *NOM - ADJ* (*maistresse inexorable*).

Non plus dans le domaine strictement métrique mais dans celui de l'éloquence, Fabri semble avoir la même intuition de sandhi externe que Palsgrave (*les deux [voyelles] se proferent ensemble*) et, comme Palsgrave, il proscriit l'élision (qu'il appelle *sinalimphe*) en fin de proposition grammaticale et/ou devant pause :

[...] mais²⁵ il fault dire de Sinalimphe en nostre vulgaire qui se fait, qu'at e feminin est en fin d'ũg terme & le prochain terme enfuyuant se cōmēce par aucun vocal ledit e feminin ne se p[ro]fere point mais les deux vocalz se proferent ensemble & des deux syllabes on en profere une comme *tu mas baiſe* ou *tu me as baiſe*.

[...] Et est a garder expreffement que lon ne face finalimphe sur le terme q[ui] fine fin d[e] sentence, pour ce que a la sentēce finee par finalimphe lon cōmenceroit & fineroit lon substance par vne syllabe, qui feroit malle proportion. Car la ou substance et parfaite proportion a pris fin lon doit pauser [...]

(FABRI 1534 [1521], vol. 2, fol. 62v-63r)²⁶

De même, Meigret et Péletier, qu'une âpre controverse publique opposa sur le domaine précis à attribuer à l'élision, insistent sur l'absence d'élision en fin de proposition :

[L'apostrophe qui note l'élision] n'a point de lieu, là ou il entrecōte quelque point autant de virgule, que de fin de clause : comme quand nous difons *vne femme bonne, apprinſe, & sage, pensera toujours de son honneur*. Autrement tu ferois vne prononciation confuse, là ou elle requiert estre distincte & faite quasi avecq quelques pofes.

(MEIGRET 1545 [1542], f2)²⁷

RONSARD (1585). *Art poetique francoys*.

24: Typographie originale, sauf italiques que nous avons rectifiés à la fin du passage.

25: Fabri parle du hiatus avant cet extrait.

FABRI (1534 [1521]). *Le Grant et vray art de pleine rethorique*.

26: Typographie de la source rendue autant que possible (la source est en gothiques); nous avons éliminé les trop nombreuses barres obliques qui ponctuent la phrase, et les italiques sont de notre fait.

MEIGRET (1545 [1542]). *Traité touchant le commun usage de l'écriture françoise*.

27: Typographie (sauf ligatures) de la source; les italiques sont de notre fait.

[...] bien fouuant s'arrête lon an lifant, fus la fin de telés diccions, ancorés que le point n'i foët pas : e an s'i arrétant, c'êt forcé de prononcer l'ê.
(PÉLETIER 1555 [1550], p. 18)²⁸

Leur controverse, si elle témoigne d'une conception différente des limites du domaine de l'élision, conforte l'idée selon laquelle celui-ci comprend plusieurs mots lexicaux. La controverse naît de la réaction de Péletier à la publication de la traduction du *Menteur* de Lucien de MEIGRET (1548), que ce dernier a fait imprimer dans son système d'orthographe phonétique, qui remplace les voyelles élidées par des apostrophes. Suite à cette publication, Péletier adresse une lettre à Meigret, qu'il publiera ensuite dans son *Dialogue* (PÉLETIER 1555 [1550]). Dans le passage qui précède la citation ci-dessus, il reproche à Meigret d'élider trop souvent les schwas finaux (Péletier note <ê> le schwa, qu'il soit élidé ou non)²⁹ :

Iè jusques ici touchè prefquê tous les poins éqez né suis d'auç toë quant a l'Escriturê, mès que premiermânt je n'oublie que fans reçon tu ôtés l'ê féminin, au lieu duquel tu mész un apostrofê a la fin des diccions, quand le mot suiuant se commance par voyelê. E ce qui le mē fêt dirê, êt que lon se peut arrêter, e bien fouuant s'arrête lon an lifant, fus la fin de telés diccions, ancorés que le point n'i foët pas : e an s'i arrétant, c'êt forcé de prononcer l'ê.
(PÉLETIER 1555 [1550], p. 18)³⁰

Meigret lui répond dans une lettre publiée en 1550, dans laquelle il déplore que Péletier soit entravé par la ponctuation traditionnelle, laquelle peut être fautive et en décalage avec à la fois le sens et la prosodie :

E qant a ce qa tu dis qa sans rézon je detourne l'e féminin par la figure de l'Apostrophe qant le mot subseqent comence par voyelle : par ce qe fur la fin de telles diccions on l'arrête bien fouuent en lizant. Ou a'tu trouué q'un lecteur se doëu'arrêter q'ès soupirs, semipozes, e poïns, e'quels l'apostrophe n'a point de pouuoer, q'il ne fasse grant tort a la graç'è mesure de la claoze : e dont bien fouuen il auient qe le fens demeure' imperfêt, obscur, e confus? Cōsidere toë memes quantes difficultés e erreurs a enjendré la maouëze punctiaçion e' langes Latin'e Gręcq.
(MEIGRET 1550a, fol. 4rv)³¹

La controverse des deux auteurs tient donc uniquement à la question de la localisation des pauses dans le discours, et à la relation entre ponctuation et pauses. De fait, Péletier et Meigret sont d'accord sur le fait qu'une pause protège toujours le schwa final de l'élision.

13.3.1.1 En somme.

L'élision que décrivent les grammairiens du xvi^e s. correspond à un domaine qui contient un ou plusieurs mots lexicaux, qui correspond parfois à une proposition grammaticale, et qui ne peut tolérer de pause en son sein. Elle est donc assez proche de l'élision métrique, mais celle-ci a moins de souplesse dans la définition de son domaine. Il nous suffit à présent d'aborder le concept de paraphonologie métrique pour expliquer

PÉLETIER (1555 [1550]). *Dialogue de l'orthographe é prononciacion françoëse*.

28: *Idem*.

29: Sur l'imprimé disponible sur Gallica (BM Lyon FC012), la barre se lit mal pour certains « ê », mais on voit toujours une trace de celle-ci : Péletier semble bien systématique dans cette notation.

PÉLETIER (1555 [1550]). *Dialogue de l'orthographe é prononciacion françoëse*.

30: Typographie (sauf ligatures) de la source.

MEIGRET (1550). *La Réponse de Louís Meigrêt a l'Apologie de Iáques Pelletier*.

31: Typographie (sauf ligatures) de la source.

comment la métrique force ce domaine à s'aligner sur les constituants métriques.

13.3.2 Paraphonologie métrique et alignement des domaines

Le terme de paraphonologie désigne, depuis KIPARSKY (1977), un ensemble de règles phonologiques qui ne s'appliquent qu'en vers, et qui s'imposent au détriment de règles phonologiques concurrentes. Dans le cas du français, DELL (2011) illustre ce concept par les diérèses qui n'existent qu'en vers ou en musique : *violette* sera prononcé [vjɔlət] en « français parisien » de conversation, mais pourra être prononcé [vijɔlət] voire [vijɔlətə] dans un vers ou dans une chanson. La paraphonologie est donc envisagée comme un ensemble de règles et/ou de contraintes plus ou moins parallèles à celles de la phonologie, mais qui, en métrique, ont préséance sur celles de la phonologie en cas de conflit.

DELL (2011). « Singing in Tashlhiyt Berber, a Language that Allows Vowel-Less Syllables ».

De fait, si le domaine de l'élision est strictement conditionné en métrique alors qu'il conserve une certaine liberté dans le discours non versifié, ce doit donc être l'effet de la paraphonologie. Nous avons avancé en sec. 13.1.3.1 que les patrons métriques français ne connaissent que deux constituants possibles : le vers et, dans les vers complexes, l'hémistiche. Si l'on voulait admettre que le mètre médiéval connaît des domaines inférieurs à l'hémistiche, reconnaissant par là la métricalité des pieds, il faudrait spécifier que la paraphonologie métrique n'assigne l'élision qu'aux domaines les plus grands qu'elle connaît, dans la mesure où générer des coupes épiques après chaque pied rendrait une quantité considérable de vers hypométriques. Si l'on s'en tient aux patrons sans constituants plus petits que l'hémistiche, il suffit en revanche de formuler le principe PARAPH.ÉLID.DOM (13.48).

(13.48) PARAPH.ÉLID.DOM.

La paraphonologie force le domaine de l'élision, quel qu'il soit, à correspondre à l'un de ses constituants.

Ce principe n'ajoute rien à la grammaire métrique ; il est redondant avec ÉLID.DOM (13.46). C'est un énoncé étimologique, qui explique d'où procède la règle ÉLID.DOM et démontre qu'elle ne tient pas que d'une logique arbitraire. Que ce soit dans nos données graphiques médiévales ou dans le discours des grammairiens du XVI^e s., l'élision a un domaine ; là où la paraphonologie métrique fait un choix arbitraire, c'est en décidant de forcer l'alignement de ce domaine avec l'un des siens. Mais le choix du domaine, en revanche, est dérivé par le faible nombre de domaines que la métrique connaît.

13.3.3 En somme

Le domaine de l'élision correspond, par l'effet de la paraphonologie métrique, à l'un des deux constituants disponibles dans le gabarit métrique : le vers ou l'hémistiche. Le domaine ainsi défini peut contenir plusieurs mots lexicaux, et coïncide non avec nos données graphiques médiévales, mais avec le discours des grammairiens du XVI^e s. Si la correspondance

entre domaine phonologique et domaine paraphonologique semble assez bonne au *xvi^e* s., il importe de s'interroger sur l'écart entre nos élisions graphiques et l'élision métrique.

13.4 Conflit de domaine

Tel qu'exposé *supra*, l'élision métrique est déterminée par un domaine lié au gabarit métrique. Ce domaine peut ne contenir qu'un seul mot lexical (premier hémistiche en 13.49 et 13.32), mais il peut aussi contenir plusieurs mots lexicaux (second hémistiche 13.49 et 13.32). Et, que les mots qu'il contienne soient lexicaux ou grammaticaux, l'élision s'applique et supprime leur voyelle finale atone devant initiale vocalique (*tere* en 13.50).

(13.49) *Vostre olifan / ne deignastes sunner*
(*Roland*, v. 1171, cf. 13.31 *supra*)

(13.50) *Tresqu'en la mer / cunquist la tere altaigne*
(*Roland*, v. 3, cf. 13.32 *supra*)

Il y a donc un écart considérable entre les élisions dont témoignent nos données graphiques, qui ne concernent guère que des mots fonctionnels et qui n'atteignent des mots lexicaux que dans quelques collocations grammaticalisées. Cet écart peut prendre sens si l'on considère, comme évoqué à la section précédente, que la métrique n'a pas de constituant plus petit que l'hémistiche, et si l'on intègre dans l'analyse le changement général de domaine computationnel qu'a connu la langue à la fin de la période de l'ancien français. Cette section commence donc par reprendre ce en quoi consiste ce changement de domaine général (sec.13.4.1), avant de s'appuyer sur cet élément pour expliquer d'une part l'écart entre les données graphiques et les données métriques et d'autre part la congruence entre les données métriques et la prosodie du *xvi^e* s. (sec. 13.4.2).

13.4.1 Élargissement du domaine computationnel

Il se produit à un moment donné de l'histoire de la langue médiévale un élargissement du domaine computationnel. Les références à ce processus dans la littérature varient, mais s'attachent la plupart du temps à le décrire sous l'angle de l'élargissement de l'unité accentuelle (MARCELLO-NIZIA 1999a, p. 172, 174, 184-189; RAINSFORD 2011b), des changements dans les paradigmes de certains pronoms et déterminants (MARCELLO-NIZIA 1999a, p. 184-189), de l'affaiblissement de l'enclise (KUKENHEIM 1971, p. 330; ADAMS 1987, p. 165; JACOBS 1993), de l'affaiblissement de l'accent de mot (POPE 1966 [1934], p. 82-83) et des consonnes finales (POPE 1966 [1934], p. 82-83; SÉGÉRAL et SCHEER 2020, GGHF, p. 406-409) ou de l'instauration de la liaison (POPE 1966 [1934], p. 82-83; SÉGÉRAL et SCHEER 2020, GGHF, p. 406-409).³² Les premières traces de sandhi consonantique entre mots lexicaux apparaissent dès *Roland* (GGHF *id.*, p. 407 : *sein Gabriel*), mais ne prennent de l'ampleur qu'à la fin de ce siècle. RAINSFORD (2011b, p. 238-241) conclut de son analyse des régularités accentuelles des textes

KUKENHEIM (1971). « Rôle de la prosodie dans l'histoire de la langue française ».

ADAMS (1987). *Old French, Null Subjects and Verb Second Phenomena*.

JACOBS (1993). « The phonology of enclisis and proclisis in Gallo-Romance and Old French. Selected Papers from LSRL XXI ».

32: Voir les synthèses de RAINSFORD (2011b, p. 3-7) et GESS, JACOBS et LAKS (2020, GGHF, p. 470-472).

versifiés que l'évolution d'un accent de mot vers un accent de groupe s'est déroulée sur une période d'un siècle à cheval sur les XII^e et XIII^e s.

Dans les témoignages des grammairiens du XVI^e s., il ne fait pas de doute que l'élision, tout comme l'accent ou la liaison consonantique,³³ concerne un domaine computationnel incluant plusieurs mots lexicaux. Les deux derniers phénomènes sont établis par la bibliographie à peu près à la même époque, mais l'élision dispose d'attestations contradictoires : la métrique l'atteste régulièrement dès *Roland*, et hors des sources métriques et musicales, aucune attestation n'advient avant le XVI^e s. Il semble raisonnable de postuler que les trois phénomènes évoluent de concert, dans un élargissement généralisé du domaine computationnel de l'ancien français. Dans ce cas, la métrique est en avance sur son temps : dès *Roland*, elle fonctionne déjà avec un domaine qui peut contenir plusieurs mots lexicaux. Ce caractère prémonitoire de l'élision métrique n'est qu'un choix par défaut, celle-ci n'ayant pas accès à d'autres constituants que le vers et l'hémistiche.

33: Voir le commentaire des *Hypomnèses* d'Henri Estienne par MORIN (2005, p. 17-18).

13.4.2 Conséquences de l'élargissement du domaine computationnel et figement graphique

Pour autant, la métrique n'est peut-être pas indifférente à l'élargissement du domaine computationnel. Si la césure épique se trouve dès les premiers textes versifiés, avec notamment *Alexis* et *Roland*, LOTE (1949-1996 [1991]) note qu'elle décline au XIV^e s. pour disparaître presque totalement à la fin du XV^e s. et être frappée d'interdiction par FABRI (1534 [1521]) et par l'anonyme *Art et Science de Rhétorique* (ANONYME ca. 1525 [1902]). Si l'on ne peut exclure que les possibilités offertes par une grammaire d'association évoluent d'elles-mêmes, par convention,³⁴ on ne peut néanmoins s'abstenir de remarquer la coïncidence, quoique légèrement décalée dans le temps : c'est peu après la fin de l'élargissement du domaine computationnel que la coupe épique disparaît. Or la coupe épique représente un petit domaine computationnel ; dans le décasyllabe le plus courant, à césure 4/6, le premier hémistiche devient trop petit pour accueillir ce qui est désormais le domaine à la fin duquel un schwa peut flotter. La disparition de la césure épique peut donc se concevoir comme un effet, à retardement, de l'élargissement du domaine computationnel, devenu trop large pour que son alignement forcé avec un hémistiche se maintienne.

34: C'est en partie le chemin que prendra la métrique à partir du XVI^e s., figeant les schwas finaux devant initiale consonantique.

L'élargissement du domaine computationnel aurait donc eu comme conséquence d'emporter la disparition de la coupe épique. Hors des paraphonologies métriques et musicales, il est selon toute vraisemblance responsable de la transition entre un état où le sandhi externe entre mots lexicaux est absent (et, où, donc, il n'y a pas lieu d'y avoir élision) à l'état du XVI^e s., où un domaine unissant plusieurs mots, lexicaux ou fonctionnels, applique une élision qui semble systématique. Or c'est précisément ce premier état dont témoignent nos données graphiques. Dans les données synthétisées au chapitre précédent, l'élision graphique n'affecte avec de hautes fréquences que des monosyllabes fonctionnels, et elle affecte parfois des mots polysyllabiques, avec des fréquences plus faibles mais encore considérables pour certains polysyllabes fonctionnels, et avec des fréquences encore plus faibles pour les mots lexicaux. Il y a donc lieu de considérer que nos données graphiques, même en écrasant

leur diachronie, représentent ce premier état de la langue, où le domaine computationnel est encore d'une taille restreinte.

Demeure, à ce stade, une question : si la graphie de la première époque, avant la fin du ^{xii} s., transcrit fidèlement l'élision, pourquoi n'admet-elle pas la transcription généralisée des nouvelles élisions qui apparaissent avec l'élargissement du domaine computationnel ? Nous n'avons pas de réponse à cette question, mais une hypothèse. Pour facile qu'elle soit, cette hypothèse n'en demeure pas moins une réelle possibilité : l'évolution des pratiques graphiques s'oppose à la notation de ces nouvelles élisions. Si l'on en croit le mètre d'*Alexis* et de *Roland*, il peut exister dès l'origine un décalage entre les élisions et leur transcription, y compris pour les monosyllabes fonctionnels (PREMAT et POGGIO [à paraître] ; MARCHELLO-NIZIA 2015). Mais celui-ci est sans commune mesure avec le décalage que prédit la simultanéité entre l'élargissement du domaine accentuel et de liaison consonantique et celui de l'élision.

L'époque à laquelle on situe ce décrochage entre la transcription de l'élision et l'élision coïncide plus ou moins avec l'accroissement de la dimension conventionnelle et logographique que (GOSSEN 1962, 1967) repère dans les graphies du ^{xiii} s. S'il ne faut pas voir dans cette dimension conventionnelle un effacement des dialectes (DEES 1985 ; LODGE 2008, p. 369-370),³⁵ il semble en revanche que le système graphique devienne de moins en moins attaché à la phonologie et à la phonétique. L'absence de transcription des élisions dues à l'élargissement du domaine computationnel pourrait donc assez naturellement découler de ce qu'elle intervient à un moment où la graphie se fait de plus en plus conventionnelle et symbolique, et ne cherche plus à représenter les alternances contextuelles. C'est peut-être ce qui justifie, trois siècles plus tard, les tentatives de Péletier et Meigret de réinventer une graphie phonétique, s'étant rendus compte que l'on écrivait trop de graphèmes ne correspondant plus à des phonèmes prononcés.³⁶

13.5 Conclusion

Dans ce chapitre, nous nous sommes intéressés à trois types de données différant des élisions graphiques étudiées jusqu'à présent : les données métriques, musicales, et celles des témoignages des grammairiens du ^{xvi} s. Après un rapide exposé du cadre théorique utilisé et de son application aux vers médiévaux (sec. 13.1), nous avons interrogé le statut de l'élision métrique, notant que celle-ci était systématique pour les polysyllabes à l'intérieur d'un hémistiche ou d'un vers, mais n'avait pas lieu en fin de vers et pas forcément lieu en fin d'hémistiche (sec. 13.2). Nous avons ensuite remarqué que le fonctionnement de l'élision métrique la rapprochait de l'élision telle que décrite, y compris hors du domaine métrique, par les grammairiens du ^{xvi} s., à ceci près que ceux-ci montrent une certaine liberté dans la définition du domaine de l'élision alors que celui-ci est strictement défini dans le vers (sec. 13.3). Cela nous a mené à interroger le hiatus observable entre d'une part nos données graphiques médiévales, et d'autre part les données métriques, musicales et grammaticales (sec. 13.4).

GOSSEN (1962). « De l'Histoire des langues écrites régionales du domaine d'oïl ».

— (1967). *Französische Skriptastudien : Untersuchungen zu den nordfranzösischen Urkundensprachen des Mittelalters*.

DEES (1985). « Dialectes et scriptae à l'époque de l'ancien français ».

LODGE (2008). « Les débuts de la standardisation du français ».

35: GOSSEN (1962, p. 262) lui-même retarde au ^{xv} s. l'expansion de la *scripta* parisienne dans de nombreuses régions d'oïl.

36: « je m'efforce de decharjer notr'écriture deç' l'èttres fuperflües, e la rendre lizable fuiuant l'uzaje de la prolaçion » (MEIGRET 1550b, fol. 3v, typographie originale).

Nous avons envisagé une hypothèse qui, pour spéculative qu'elle soit, autorise à entrevoir la cohérence du système phonologique et permet de proposer une interpolation là où les données sont muettes. Cette hypothèse, qui sera conservée dans l'analyse phonologique proposée dans le chapitre suivant, prédit que l'élargissement du domaine de l'élision est commun avec celui du domaine de l'accentuation et du traitement des consonnes finales. Dans cette hypothèse, l'élargissement du domaine computationnel a donc dû s'opérer entre la fin du *xii^e* s. et la moitié du *xiii^e* s. (sec. 13.4.1). Sur ce fondement, la théorie prédit que l'élision phonologique cesse de correspondre à l'élision graphique pendant cette période, mais que ce changement dans le domaine de l'élision ne donne pas lieu à une adaptation des pratiques graphiques, qui continuent de noter conventionnellement les schwas finaux des polysyllabes indépendamment de leur élision contextuelle (sec. 13.4.2).

14.1 Introduction

Ce chapitre vise à proposer une analyse phonologique des élisions relevées d'une part dans nos données graphiques et d'autre part dans les données métriques, musicales et dans les témoignages des grammairiens du *xvi*^e s. Nous considérons que les données graphiques témoignent de la première période de l'a.fr., où le domaine computationnel ne peut concerner qu'un seul mot lexical, tandis que les données métriques et musicales et, postérieurement, le discours des grammairiens de la Renaissance, témoignent de la deuxième période de l'a.fr., lorsque le domaine computationnel s'est élargi et peut contenir plusieurs mots lexicaux. Ces périodes 1 et 2 correspondent donc à l'empan du domaine computationnel, et non aux *Period I* et *Period II* de POPE (1966 [1934]).¹

Après avoir présenté un rapide résumé de l'ensemble des données que notre analyse doit prendre en compte (sec. 14.1.1), nous détaillons dans un premier temps les conditions phonologiques de l'élision, en proposant de représenter la capacité d'une voyelle à s'élider par le flottement de sa mélodie (sec. 14.2). Nous distinguons deux types de flottements : le premier, qui vient remplacer le diacritique d'élision utilisé au chap. 12, fait la différence entre les monosyllabes fonctionnels qui s'élident et ceux qui résistent à l'élision (sec. 14.2.1); le second atteint tous les schwas finaux (sec. 14.2.2). Nous détaillons ensuite comment le caractère flottant de la mélodie vocalique finale interagit avec l'environnement au sein d'un domaine, de manière à ne dériver l'élision que devant initiale vocalique (sec. 14.2.3). Au terme de cette section, nous disposons d'une représentation des voyelles sujettes à l'élision et d'un mécanisme d'élision.

Dans un second temps, nous envisageons le domaine dans lequel s'applique ce mécanisme, avec pour objectif de rendre compte de la différence entre l'élision de la période 1 et celle de la période 2 sans toucher au mécanisme phonologique (sec. 14.3). Nous proposons de définir l'élision de la période 1 comme étant un phénomène lié à la cliticité (sec. 14.3.1), et proposons de considérer que la cliticité, dans la période 1 de l'a.fr., dépend d'une condition syntaxique (sec. 14.3.1.1) mais aussi d'une condition de taille du clitique (sec. 14.3.1.2). Par contraste, l'élision de la période 2 n'est plus restreinte aux clitics mais s'applique dans un domaine plus large (sec. 14.3.2). Nous envisageons la possibilité de définir ce domaine comme étant le syntagme phonologique de la hiérarchie prosodique, mais concluons que la labilité de celui-ci et l'absence de données directes ne nous permettent pas de tirer de conclusions satisfaisantes sur ce point (sec. 14.3.2.1). En revanche, nous pouvons observer deux contextes dans lesquels nos données témoignent de l'élargissement du domaine : d'une part le contexte *ADJ + NOM* et d'autre part les contextes d'élision progressive du pronom sujet *jo* (sec. 14.3.2.2). Enfin, nous montrons que le mécanisme phonologique développé pour la période 1 s'applique tel

1: Pour POPE (1966 [1934], p. 9, 23), *Period I* correspond à « Gallo-Roman and Early Old French », et *Period II* à « Later Old and Middle French ».

quel à la période 2 ; la seule différence entre ces deux périodes étant le domaine de l'élision (sec. 14.3.3).

14.1.1 Résumé des données

Dans cette section, nous résumons les données avancées précédemment pour l'élision graphique (période 1) et pour l'élision métrique et celle du xvi^e s. (période 2).

14.1.1.1 Période 1

La figure 14.1 reproduit le graphe présenté en sec. 12.2.2, et représente l'ensemble des données graphiques sur lesquelles nous avons travaillé. Elle correspond donc à la période 1 de l'a.fr. Dans cet état de langue, l'élision systématique ne concerne que des monosyllabes fonctionnels ;

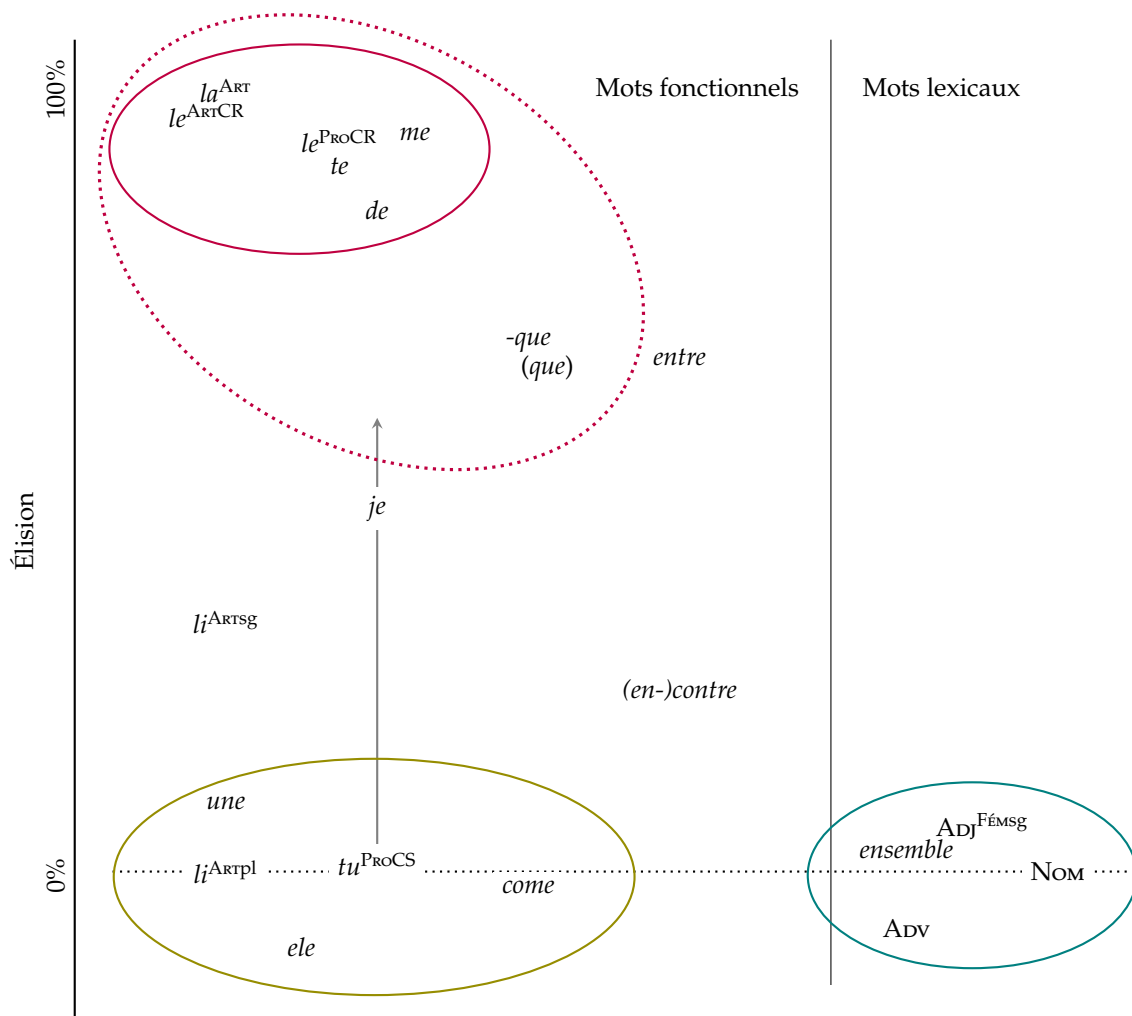


FIGURE 14.1 – Reproduction à l'identique de la fig. 12.6. Distribution des taux d'élision en fonction des des catégories et items. Un taux d'élision négatif signifie qu'on trouve plus de défauts devant #C que devant #V. Les valeurs des ordonnées sont approximatives.

l'élision optionnelle concerne des monosyllabes fonctionnels et des polysyllabes fonctionnels. L'élision est très faible pour un monosyllabe fonctionnel (*ele* 6% de défauts devant #V mais 14% devant #C, le tout avec une forte variation diatopique)² et pour certains mots lexicaux (*ensemble*, adjectifs féminins). L'élision, même optionnelle, ne concerne pas tous les monosyllabes fonctionnels : elle ne concerne ni *li* pl ni *tu*. En sec. 12.3.1 *supra*, nous avons réduit cette typologie à un seul critère phonologique, hérité du proto-a.fr. : COND CV. La plupart des mots qui répondaient à cette condition en proto-a.fr. sont, en a.fr., élidés à plus de 90%,³ et correspondent à l'ellipse violette en trait plein en fig. 14.1. Le pronom personnel *jo*, *je* est un cas particulier, dans la mesure où ses domaines syntaxiques ne lui permettent que progressivement de s'élider, comme indiqué par la flèche qui représente son évolution diachronique, de 0 à ~60% d'élision dans notre corpus (cf. sec. 14.3.2.2 *infra*). Enfin, l'élément *que* et le *que* des locutions en *base* + *-que* montrent un taux d'élision légèrement plus faible, ce qui les fait apparaître comme des cas particuliers au sein de l'ensemble des monosyllabes de structure CV dont l'élision est systématique (ellipse violette en pointillés ; cf. sec. 14.2.3.5 *infra*).

2: Cf. fig. 9.25 en sec. 9.2.8.1

3: Pour *de*, l'exception anglo-normande est ici ignorée (moy. a.n. = 59%).

En revanche, les mots fonctionnels qui ne répondaient pas à la condition COND CV en proto-a.fr. ne connaissent pas l'élision (ellipse ocre). C'est le cas de l'article pluriel *li* et du pronom sujet *tu* — pour lesquels nous reconstruisons des formes diphtonguées en proto-a.fr. (*li* pl ← /lij/ ← ɪllī; *tu* ← /tuw/ ← tū) — ainsi que de *ele* et *une*. Enfin, les mots lexicaux échouent toujours à COND CV, et ne s'élident donc pas, à moins qu'ils ne perdent leur autonomie dans les collocations où en devenant des préfixes (ellipse bleue).

Quant aux mots n'appartenant à aucun cluster en fig. 14.1, ils sont de deux types : d'une part l'article singulier *li*, qui connaît une certaine tendance à l'élision (33%), et d'autre part les prépositions *entre*, *contre* et *encontre*. Pour *li* sg, nous avons expliqué que sa formation analogique semblait avoir copié la forme de *li* pl mais pas son statut vis-à-vis de l'élision : *li* sg est bien élidé dans 33% de ses occurrences, et ce principalement au Sud-Ouest du domaine. Enfin, pour *entre*, *contre* et *encontre*, nous avons noté que leurs défauts de <e> étaient liés à des constructions au moins partiellement grammaticalisées ; dans ce cadre, leur tendance à devenir un préfixe explique leur élision puisque, devenant préfixes, ils cessent d'être des mots autonomes.

14.1.1.2 Période 2

Comme exposé au chapitre précédent, les données dont nous disposons pour la période 2 sont essentiellement indirectes. Nos déductions quant à l'élision de cette période se fondent sur la rencontre de trois éléments concordants : le traitement de l'élision en métrique et en musication, le discours des grammairiens au xvi^e s. et la communauté de destin phonologique entre l'élision, le statut des consonnes finales et l'évolution de l'unité accentuelle. Ces éléments, pris ensemble, permettent de supposer raisonnablement que le domaine de l'élision peut contenir, dans cette seconde période, plusieurs mots lexicaux ainsi que des mots fonctionnels polysyllabiques. En revanche, nous n'avons aucune donnée indiquant que *li* pl ou *tu* sont atteints par l'élision. Pour le premier, l'affaiblissement

puis la disparition du système bicasuel le laisse sans descendance. Pour le second, il est probable qu'il a toujours conservé une résistance au moins partielle à l'élision ; encore aujourd'hui, l'élision de *tu* pronom sujet est un trait stylistiquement et sociolinguistiquement marqué.

Aussi, le panorama général de la seconde période est le suivant : le domaine de l'élision s'élargit, et celle-ci n'est plus restreinte aux seuls clitiques. L'élision n'est donc plus un phénomène qui n'atteint que des mots fonctionnels, et le sandhi externe se libère du lien entre les clitiques et leur hôte. Si nous serons amenés, en sec. 14.3.2.2, à avancer quelques éléments de détail sur l'élargissement du domaine computationnel, celui-ci reste majoritairement hors champ : nous nous intéressons avant tout à l'élision en tant que processus au sein d'un domaine.

14.1.2 Cahier des charges général

Notre analyse se fixe donc les buts suivants. Elle doit donc rendre compte de deux états diachroniques différents : l'un (période 1) où l'élision ne concerne en premier lieu que certains monosyllabes fonctionnels et, ponctuellement, quelques polysyllabes, et l'autre (période 2) où elle se généralise à toutes les finales atones. Au sein de la période 1, l'analyse doit permettre la distinction, si possible sans recourir à un diacritique, entre les monosyllabes sujets à l'élision (*le, la, me, te* CR etc.) et ceux qui y résistent (*tu* CS, *li* PL, et tous les monosyllabes lexicaux). En période 2, l'analyse doit permettre la généralisation de l'élision à tous les polysyllabes, mais sans contaminer les monosyllabes fonctionnels qui résistent à l'élision (*li* pl, *tu* CS), pour lesquels aucune donnée disponible ne témoigne de l'élision.

Dans ce qui suit, nous proposons de distribuer les rôles à différents outils, avec notamment une division du travail entre phonologie et syntaxe. Du côté de la phonologie, nous montrons que le trait diacritique qui sépare les monosyllabes qui s'élident des autres monosyllabes peut être efficacement transcrit sous la forme du flottement de la mélodie vocalique finale (sec. 14.2). Nous avançons ensuite que l'adoption d'un simple mécanisme issu de la phonologie de gouvernement, la suppression des VC vides, permet de dériver l'élision à partir du flottement mélodique, sans avoir besoin de recourir à une contrainte anti-hiatique (14.2.3).

Du côté de la syntaxe (sec. 14.3), et de son interface avec la phonologie, nous remarquons que la différence entre les élisions de la période 1 et celles de la période 2 tient principalement à ce que l'élision de la période 1 ne concerne que des clitiques, mais ne concerne pas tous les monosyllabes fonctionnels. Nous proposons donc une définition restrictive de la cliticité, qui parvient à isoler les mots fonctionnels qui ne s'élident pas de ceux qui s'élident. Pour ce faire, nous définissons la cliticité non seulement en fonction de la proximité syntaxique entre le clitique et son mot tête (sec. 14.3.1.1) mais également en fonction de la taille de l'objet clitique (14.3.1.2).

14.2 Condition phonologique : le flottement

L'une des difficultés principales de l'analyse repose sur le fait qu'il faut donner une représentation phonologique au diacritique qui sépare les monosyllabes fonctionnels *li* pl et *tu* CS des autres monosyllabes fonctionnels. En effet, tel qu'avancé en chap. 12, si la plupart des monosyllabes semblent être atteints par une élision systématique (< 90% d'élision), *li* pl et *tu* CS ne s'élident jamais. Cette difficulté est d'autant plus grande que les monosyllabes qui s'élident ne se terminent pas tous en schwa (*la* ART,⁴ *li* sg), et que la dissociation entre *li* pl et *li* sg (qui s'élide facultativement) interdit de fonder cette distinction sur l'identité mélodique des voyelles. Ces difficultés supplémentaires rendent également caduque l'hypothèse d'un mécanisme général de réduction des hiatus, dans la mesure où celui-ci devrait être en mesure de faire la distinction entre *li* pl et *li* sg, et entre *tu* CS et *jo* CS; dans ces deux couples de distinctions, la structure syntaxique est nécessairement identique — on ne peut imaginer une structure syntaxique différente pour *li* pl et pour les autres articles, ni pour *tu* CS et pour les autres pronoms personnels sujets.

Jusqu'à présent, nous avons représenté l'identité de nos monosyllabes fonctionnels sous la forme d'une matrice comportant un attribut de contenu mélodique et un attribut d'élision. Nous avons également avancé que, pour les deux monosyllabes fonctionnels résistants à l'élision que l'on trouve dans nos données, la valeur négative de l'attribut d'élision (symbolisée ×) était héritée d'un test phonologique, baptisé COND CV, qui avait décidé de la capacité du clitique à s'élider en fonction de sa correspondance à un simple gabarit CV. En sec. 12.3.1, ce test a été formulé de la manière suivante :

- (14.1) COND CV
Un mot n'est candidat à l'élision graphique que s'il se réduit à une structure *consonne* + *voyelle* atone.

Dans les matrices utilisées au chapitre 12, le trait issu de COND CV était donc représenté comme un attribut distinct de la mélodie, puisque pour que COND CV fasse la distinction entre *li* pl et *li* sg et entre *tu* CS et *jo* CS, il fallait rejeter ce test dans la préhistoire du français, époque à laquelle nous pouvions reconstruire des formes diphtonguées /lij/ et /tuw/ pour *li* pl et *tu* CS, mais pour aucun des clitiques sujets à l'élisions. Nous pouvions donc représenter nos monosyllabes fonctionnels sous la forme des matrices suivantes en (14.2 - 14.4).

4: Et probablement aussi le pronom *la*; sur l'absence de données pour le pronom régime *la* mais la plausibilité qu'il se comporte comme *le*, cf. sec. 9.2.1.

(14.2) Catégorie à élision systématique

(a)	$\left[\begin{array}{l} \text{Mélodie} = x \\ \text{Élision} = \checkmark \end{array} \right] \leftarrow \text{COND}_{\text{CV}} = \checkmark$	
(b) <i>la</i>	$\left[\begin{array}{l} \text{Mélodie} = /la/ \\ \text{Élision} = \checkmark \end{array} \right] \leftarrow \begin{array}{l} /la/ \quad \neg \\ \text{COND}_{\text{CV}} = \checkmark \end{array}$	
(c) <i>jo</i>	$\left[\begin{array}{l} \text{Mélodie} = /ʒo/ \\ \text{Élision} = \checkmark \end{array} \right] \leftarrow \begin{array}{l} /ʒo/ \quad \neg \\ \text{COND}_{\text{CV}} = \checkmark \end{array}$	

x : contenu mélodique indifférent.

\neg : la représentation du proto-a.fr. décide de la valeur de l'attribut de COND_{CV} .

(14.3) Catégorie à élision facultative de type CV : *li* sg

(a) <i>li</i> sg	$\left[\begin{array}{l} \text{Mélodie} = /li/ \\ \text{Élision} = \checkmark \end{array} \right] \leftarrow \begin{array}{l} /li/ \text{ pl} \\ \text{COND}_{\text{CV}} = \checkmark \end{array} \leftarrow \text{le, la}$	
------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

\leftarrow : formation analogique, cf. sec. 8.1.1.

(14.4) Catégorie à élision prohibée

(a)	$\left[\begin{array}{l} \text{Mélodie} = x \\ \text{Élision} = \times \end{array} \right] \leftarrow \text{COND}_{\text{CV}} = \checkmark$	
(b) <i>li</i> pl	$\left[\begin{array}{l} \text{Mélodie} = /lij/ \\ \text{Élision} = \times \end{array} \right] \leftarrow \begin{array}{l} /lij/ \text{ pl} \quad \neg \\ \text{COND}_{\text{CV}} = \times \end{array}$	
(c) <i>tu</i> CS	$\left[\begin{array}{l} \text{Mélodie} = /ty/ \\ \text{Élision} = \times \end{array} \right] \leftarrow \text{COND}_{\text{CV}} = \times$	

Ces matrices permettent d'expliciter la dissociation entre le contenu mélodique du monosyllabe fonctionnel et son attribut de capacité à s'élider. Elles mettent également en exergue le fait que cet attribut de capacité à l'élision est hérité d'un état précédent de la langue, ce qui permet sa désolidarisation vis-à-vis de la forme synchronique du mot, ce qu'illustre la différence entre *li* sg (14.3) et *li* pl ((b)), et qui ne peut non plus provenir de sa classe grammaticale, comme l'illustre la différence entre l'article *la* ((b)) et *li* pl ((b)), et entre le pronom personnel sujet *jo* ((c)) et le pronom personnel sujet *tu* ((c)). En revanche, ces matrices ne sont - et ne doivent pas être - des représentations phonologiques. Ce ne sont que des outils destinés au phonologue, et le trait d'élision n'est ici qu'un diacritique. Il convient désormais de trouver quelle forme peut prendre ce diacritique en termes de représentation phonologique.

14.2.1 Flottement diacritique et diachronique

14.2.1.1 Transcription du diacritique par le flottement

Nous avons donc besoin d'un outil qui permettent de transcrire en phonologie le diacritique de capacité à l'élision, et ce diacritique doit prendre une forme qui ne tient pas à la mélodie (en synchronie, *li* sg et *li* pl sont *a priori* homophones) et n'est pas hérité de la syntaxe (*jo* et *tu* ne peuvent avoir une position syntaxique différente). La théorie autosegmentale (GOLDSMITH 1972) propose un outil qui permet l'encodage

GOLDSMITH (1972). *Autosegmental Phonology*.

phonologique de la distinction que nous cherchons à établir. En effet, dans la phonologie autosegmentale, la représentation phonologique n'est pas réductible à une simple succession de mélodies, mais doit être divisée en plusieurs plans connectés entre eux. Dans les représentations autosegmentales, rien n'impose que chaque élément d'un plan soit strictement associé à un élément dans chaque autre plan et, parmi ces plans se trouvent le plan positionnel et le plan mélodique. Cette théorie permet donc d'envisager l'existence de mélodies sans position et de positions sans mélodie. On peut donc envisager, parmi les situations possibles (BENDJABALLAH et HAIDEN 2008), l'association entre une position x et une mélodie y (14.5 a), mais également la présence d'une position sans mélodie (14.5 b), d'une mélodie sans position (14.5 c), ou la présence d'une mélodie et d'une position sans lien entre elles (14.5 d).

(14.5) Configurations entre plan positionnel et mélodique



Les différents types d'association en (14.5 b - d) représentent des situations marquées vis-à-vis de la situation par défaut, où mélodie et position sont associées entre elles dans la représentation lexicale (14.5 a). BENDJABALLAH et HAIDEN (2008, p. 28-29) qualifient (14.5 b) de site vide (en. *empty site*) et (14.5 c) de segment flottant (en. *floating/epenthetic segment*). Ces configurations sont celles proposées par KAYE (1990) et SPENCER (1986) pour le site vide et RUBACH (1986) et KENSTOWICZ et RUBACH (1987) pour la mélodie flottante, dans la représentation des alternances V/Ø interconsonantiques. La représentation 14.5 d est celle proposée par SCHEER (1997, 1998).

Dans la phonologie du français, le concept de mélodie flottante (14.5 d) a également été utilisé pour la description des consonnes de liaison et de l'élision des monosyllabes fonctionnels (CLEMENTS et S. KEYSER 1983; ENCREVÉ 1988; TRANEL 1987, 1995b,a, 1996, 2000). Le flottement de la mélodie, ou latence, permet de définir un statut spécifique pour les consonnes de liaison et les voyelles élidables, qui les rend plus propices à la suppression que les autres mélodies (qui sont, elles, associées à une position comme en 14.5 a *supra*) mais qui ne les supprime pas de la représentation lexicale de manière à pouvoir les mobiliser dans certains contextes syntaxiques et phonologiques. Notons que, dans le cas de l'élision du français moderne décrit par Tranel, l'élision ne concerne que des monosyllabes fonctionnels se terminant par un schwa (*le ART et PRO, je, me, te, se, ce PRO, de, ne, que*) et *la* (*ART et PRO*), pour lequel TRANEL (2000, p. 59, 60) propose un supplétisme (il remplace /la/ par /le/ du masculin devant initiale vocalique), a mécanisme adopté par ailleurs pour décrire les alternances C/Ø des consonnes de liaison.

La proposition de Tranel permet donc de transcrire notre diacritique d'élision au sein de la représentation phonologique : toute voyelle flottante peut être sujette à élision, et ne peuvent être sujettes à élision que les voyelles flottantes (14.6). Nous pouvons donc remplacer notre diacritique d'élision par le caractère flottant de la voyelle (14.7 - 14.8).

BENDJABALLAH et al. (2008). « A Typology of Emptiness in Templates ».

KAYE (1990). « 'Coda' licensing ».

SPENCER (1986). « A non-linear analysis of vowel-zero alternations in Polish ».

RUBACH (1986). « Abstract vowels in three dimension phonology: the yers ».

KENSTOWICZ et al. (1987). « The Phonology of Syllabic Nuclei in Slovak ».

SCHEER (1997). « Vowel-zero alternations and their support for a theory of consonantal interaction ».

— (1998). « Governing domains are head-final ».

CLEMENTS et S. KEYSER (1983). *CV Phonology. A Generative Theory of the Syllable*.

ENCREVÉ (1988). *La liaison avec et sans enchaînement. Phonologie tridimensionnelle et usages du français*.

TRANEL (1987). *The Sounds of French: An introduction*.

— (1995). « The representation of French final consonants and related issues ».

— (1995). « Current issues in French phonology: Liaison and position theories ».

— (1996). « French liaison and elision revisited: A unified account within Optimality Theory ».

— (2000). « Aspects de la phonologie du français et la théorie de l'optimalité ».

- (14.6) ÉLID.FLOT.
Les voyelles flottantes sont candidates à l'élision, les voyelles non flottantes ne sont pas candidates à l'élision.
- (14.7) ÉLISION = ✓ ⇒ V flottante.
- (14.8) ÉLISION = × ⇒ V non flottante.

Si l'on admet (14.6), il faut donc postuler que les monosyllabes fonctionnels sujets à élision ont une voyelle flottante dans leur représentation lexicale, tandis que les monosyllabes fonctionnels qui résistent à l'élision ont une voyelle associée à une position dans leur représentation lexicale. On postule donc les oppositions entre *le* et *li pl* et *jo* et *tu* telles qu'en (14.9).

- (14.9) Voyelles flottantes en synchronie

(a) ×	(b) × ×	(c) ×	(d) × ×
l ə	l i	j o	t u

14.2.1.2 Origine du flottement diacritique

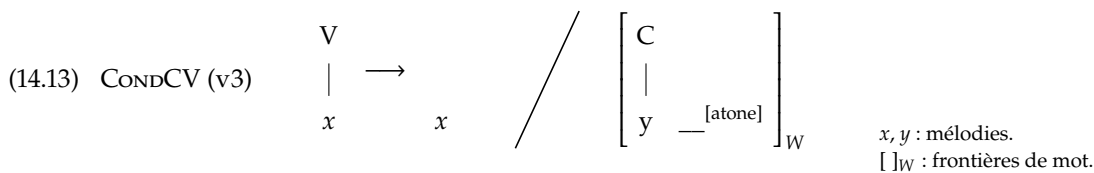
Nous avons avancé en sec. 12.3.1.2 que la résistance de *li pl* et *tu CS* à l'élision n'était pas nécessairement arbitraire et pouvait avoir une causalité historique. Ce sont les deux seuls monosyllabes fonctionnels dont l'étymon latin se termine par une fermée longue, et il est probable que ces voyelles n'ont pas échappé à la diphtongaison mais que celle-ci n'a pas laissé de trace visible. Nous reconstruisons donc des formes /lij/ et /tuw/ pour ces mots en proto-a.fr.

De fait, ces formes à approximante finale classent *li pl* et *tu CS* parmi les monosyllabes fonctionnels à consonne finale : comme *les* /ləs/, *des* /dəs/, ce mots ne s'élident pas. Nous proposons donc, en sec. 12.3.1, un mécanisme COND CV qui déciderait de la capacité du monosyllabe à s'élider en fonction de sa conformité à une structure *consonne - voyelle* (CV). Nous pouvons désormais reformuler COND CV de manière à ce que ce test ne génère plus un diacritique mais à ce qu'il aboutisse à la dissociation entre position et mélodie des monosyllabes fonctionnels CV (14.10).

- (14.10) COND CV (v3)
Toute voyelle finale atone d'un mot de structure CV devient flottante.

Cette condition cible directement le caractère consonantique ou vocalique des éléments constitutifs de la structure du mot; ce caractère n'est pas spécifié dans le plan positionnel en (14.9), qui n'est composé que de positions ×. Plutôt que d'introduire un niveau de représentation supplémentaire comme en (14.11), nous remplaçons directement les positions × par des positions étiquetées C et V (14.12).

Cette modification dans l'étiquetage de nos positions permet la formulation synthétique de COND CV dans le formalisme habituel des dérivations en (14.13), où *x* et *y* représentent des mélodies.



La règle COND CV sert donc à transcrire en information phonologique le diacritique de capacité d'élision que nous représentions dans les matrices précédemment. Cette transcription se fait sous la forme du flottement de la voyelle finale atone, pour les monosyllabes qui sont atteints par l'élision. Si l'on veut conserver cette fonction de transcription de diacritique, il faut considérer que COND CV était une règle active avant l'a.fr., mais qu'elle n'est plus active en a.fr. C'est en rejetant COND CV en proto-a.fr. que l'on peut expliquer que *li pl* et *tu CS* ne s'élident pas (leurs formes proto-a.fr. /lij/ et /tuj/ ne correspondent pas à une structure CV) tandis que les autres monosyllabes atones s'élident, y compris *li sg.*, formé par analogie sur la mélodie de *li pl* mais dont les lignes d'associations proviennent visiblement des autres articles définis et déterminants. Si la voyelle flottante de *li sg* vient bien de la pression analogique des autres articles et pronoms en /l^V/, alors l'analogie rejoue, en synchronie, l'effet de COND CV. Mais la règle, en tant que telle, est inactive depuis le début de l'a.fr. Ce scénario de formation de *li sg* est synthétisé en (14.14).

(14.14) Formation de *li sg*

- (a) Les mélodies de *li sg* proviennent d'une analogie avec *li pl* /li/ (issu de la monophthongaison de /lij/)
- (b) Les lignes d'association de *li sg* proviennent d'une analogie avec les autres ART DÉF et PRO de structure CV
- (c) Donc *li sg* /li/ a une structure CV et celle-ci reproduit ce qu'aurait été l'effet de COND CV. *Li sg* (d) diffère donc de *li pl* (e) uniquement par sa mélodie flottante.



14.2.1.3 En somme

La dissociation entre la mélodie synchronique des monosyllabes fonctionnels et leur capacité à s'élider peut être représentée par le flottement des voyelles finales de ceux qui s'élident. Ce flottement est hérité du proto-a.fr., où la règle COND_{CV} a entraîné le flottement de la voyelle finale des monosyllabes atones de structure CV et n'a pas atteint la voyelle finale des mots dont la structure n'est pas CV atone. Nous obtenons ainsi la distinction entre les monosyllabes fonctionnels qui s'élident et ceux qui ne s'élident pas. Ces mots sont ceux dont l'élision est transcrite dans la graphie, à l'exception des quelques cas de collocation.

14.2.2 Flottement non diacritique et synchronique

Le flottement diacritique qui atteint les monosyllabes fonctionnels suffit pour prédire l'élision graphique, qui est celle de la période 1, mais il ne prend pas en compte celle de la période 2. En période 2, comme l'attestent la métrique et le témoignage des grammairiens, tout schwa final s'élide. Il y a plus : en période 1, certains polysyllabes sont capables de s'élider dans certains cas de collocation, et cette mécanique ne peut être réduite à celle de la réduction du hiatus interne.

De manière à garder un seul mécanisme général d'élision, l'élision de la période 2 doit, comme celle de la période 1, dépendre du statut flottant de la voyelle finale — cela permet aussi de protéger *li pl* et *tu* de l'élision en période 2. En effet, si *li pl* va disparaître en période 2 avec l'affaiblissement du système casuel, il n'en va pas de même de *tu* et, dans notre corpus, ces deux items résistent toujours à l'élision. Il s'agit donc d'ajouter une seconde règle de dissociation. Plusieurs formulations de cette règle sont envisageables ; il faut qu'elle s'applique aux paroxytons lexicaux et aux polysyllabes fonctionnels, et qu'elle n'atteigne pas *li sg* et *tu CS*. Du fait que seuls les monosyllabes fonctionnels disposent de voyelles finales atones autres que schwa, nous avons adopté la règle COND._ə# en (14.15). En plus d'atteindre sa cible (les schwas finaux des polysyllabes ; 14.16 a), cette règle n'atteint pas les monosyllabes lexicaux (dont la voyelle est toujours autre que schwa ; 14.16 b). Quant aux monosyllabes fonctionnels en schwa, peu importe qu'ils soient dans les cibles de COND._ə : leur voyelle finale est déjà flottante s'il s'agit d'un schwa (14.16 c), et n'est pas un schwa si elle n'est pas flottante (*li*, *tu*).

(14.15) COND._ə#
Tout schwa final devient flottant.

(14.16) Par définition, COND._ə#...

- (a) S'applique sur : *ensemble* (/enseibl^ə/), *entre* (/entr^ə/), *une* (/yn^ə/), *amie* (/ami^ə/), etc.
- (b) Ne s'applique pas sur : *li pl*, *tu CS* et les monosyllabes lexicaux
- (c) Est redondant sur : *je* (/ʒ^ə/), *me*, *te*, *de*, *que*

Il n'est pas nécessaire de spécifier que COND._ə# n'est pas une règle active en synchronie en a.fr. ; l'application continue de cette règle ne génère aucune mauvaise prédiction, puisque tout schwa final est sujet à l'élision.

Il y a même un certain nombre de bénéfiques à garder COND.ə# active en synchronie. D'une part, ses cibles contiennent les mots fonctionnels, qui représentent une classe ouverte : garder la règle active permet de rendre flottant le schwa final de tous les mots qui intégreraient le lexique après la période préhistorique. D'autre part, ce processus de dissociation des schwas finaux trouve un écho, en français médiéval, dans celui des schwas internes placés en hiatus ou entre deux consonnes dont l'une est nasale ou sonore.

14.2.2.1 Les hiatus et schwas internes

Nous avons vu, en chap. 11, que la réduction des hiatus internes et la syncope étaient deux processus actifs et en expansion en a.fr. La réduction des hiatus internes, non attestée dans *Roland* (SHORT 2007, p. 92), est attestée dès le second tiers du XII^e s., mais nous avons montré que certains lemmes n'étaient atteints que plus d'un siècle plus tard (sec. 11.3.1.2). La syncope, quant à elle, est attestée dès *Alexis* (SHORT 2007, p. 95), mais elle ne se répand de lemme en lemme que progressivement, à peu près au même rythme que la réduction des hiatus. Si le processus qui vient affaiblir les schwas internes pour générer la réduction des hiatus et la syncope prend plus d'un siècle pour atteindre l'ensemble des lemmes présentant un contexte phonologique favorable, c'est donc que la règle est active en synchronie en a.fr.

Or le processus qui affaiblit les schwas internes peut procéder de la même logique générale que celui qui affaiblit les schwas finaux. Nous pouvons donc être amenés à postuler que les schwas internes en hiatus ou entre deux consonnes (dont l'une est liquide ou nasale) subissent la même dissociation avec leur position V que les schwas finaux. La seule différence, ici, réside dans la chronologie : nous n'avons pas de trace de 'retard' de la dissociation en finale, alors que le hiatus et la syncope se déclenchent à date variable selon le lemme. Mais, puisque nos sources sur l'élision en période 2 sont essentiellement indirectes et que la période 2 advient alors que la plupart des lemmes à hiatus interne ou syncope semblent déjà atteints, il est probable que si de tels 'retards' existaient en finale, nous ne pourrions pas les percevoir.

Nous pouvons donc reformuler COND.ə# (14.15) comme étant une règle de dissociation plus générale, active en synchronie en ancien français (14.15). Le détail de sa différente vitesse d'application et/ou de son caractère facultatif en position interne est laissé hors-champ, n'étant pas le sujet de ce travail. Le traitement de la réduction des hiatus est abordé en sec. 14.2.3.4.

(14.17) COND.ə
Tout schwa est devient flottant.

14.2.2.2 En somme

Nous proposons donc de considérer que l'élision n'a lieu que pour une voyelle flottante, et que deux règles peuvent être responsables du flottement d'une voyelle. La première, COND.CV, ne concerne que les mots atones dont la structure est CV, est indifférente à la nature de la mélodie

qu'elle rend flottante, et est active en proto-a.fr. De ce point de vue, l'a.fr. ne fait qu'hériter les flottements décidés par COND.CV en proto-a.fr. La seconde, COND.ə, est active en a.fr., n'a pas de contrainte de structure ou d'environnement, mais ne cible que les schwas.

14.2.3 Mécanique phonologique

Cette section vise à déployer le mécanisme par lequel nous entendons rendre compte de l'élision. Elle s'appuie en premier lieu sur différentes propositions faites par Bernard Tranel, que nous discutons afin d'en tirer les outils qui seront utiles à notre analyse. La situation de l'a.fr. étant différente de celle du fr.mod. analysée par Tranel, nous ne parviendrons pas aux mêmes conclusions formelles que lui. Après une introduction aux analyses proposées par TRANEL (1996, 2000) (sec. 14.2.3.1), nous reprenons chacune de ces deux analyses pour en mesurer l'adéquation avec nos données (sec. 14.2.3.2 et 14.2.3.3), avant de mettre au point notre propre analyse à partir d'éléments issus de Tranel et d'éléments issus de la phonologie autosegmentale et de la phonologie de gouvernement (14.2.3.4).

Si, dans la description des analyses de Tranel, nous empruntons le vocabulaire de la Théorie de l'Optimalité (PRINCE et SMOLENSKY 2004) que l'auteur utilise, il importe de signaler que notre analyse n'adopte pas ce cadre théorique. Les différents processus et les différentes règles que nous définissons ne sont donc pas des contraintes au sens de cette théorie. Elles pourraient le devenir si l'on voulait utiliser cette théorie de la computation, mais ce n'est pas notre but ici.

14.2.3.1 L'analyse de B. Tranel : contrainte anti-hiatique, ou contrainte anti-mélodie flottante

L'élision, par définition, ne se produit qu'en hiatus, et le hiatus est une configuration marquée. C'est donc par le biais d'une contrainte anti-hiatique, *VV, que TRANEL (2000) provoque l'élision de *je, me, te, se, ce, de, ne, et que*, recourant à un supplétisme pour *la* (les élisions de *la* « résult[en]t en fait de l'utilisation du masculin *le* à des fins phonologiques prioritaires. » TRANEL 2000, p. 59). Mais, dans cette analyse anti-hiatique, Tranel ne fait jouer aucun rôle au flottement du schwa élidé : il génère l'élision en plaçant sous *VV la contrainte responsable d'assurer la réalisation des schwas, comme dans la hiérarchie en (14.18).⁵ Cette hiérarchie de contrainte peut être glosée ainsi : supprimer une voyelle (autre que schwa) est pire que de tolérer un hiatus, et tolérer un hiatus est pire que de supprimer un schwa voyelle.

(14.18) MAX(V) ≫ *VV ≫ MAX(schwa) (TRANEL 2000)

Dans cet article, si Tranel reconnaît qu'il faut donner au schwa un statut plus « faible » que celui des autres voyelles, la possibilité de considérer le schwa comme une voyelle flottante n'est mentionnée qu'en note (p. 59, n. 29) et ne joue aucun rôle dans l'analyse. C'est dans un article précédent (TRANEL 1996) que l'auteur utilise le caractère flottant du schwa pour générer l'élision. Dans cet article, Tranel introduit une contrainte qui milite contre l'insertion de l'élément structurel qui permettrait à une

TRANEL (1996). « French liaison and elision revisited: A unified account within Optimality Theory ».

TRANEL (2000). « Aspects de la phonologie du français et la théorie de l'optimalité ».

PRINCE et al. (2004). *Optimality Theory : Constraint interaction in generative grammar*.

5: La hiérarchie en (14.18) est une version simplifiée de la hiérarchie proposée par Tranel, qui distingue les voyelles selon leur ouverture de manière à générer des élisions de voyelles fermées en français « familier » (*le taxi qui arrive* [loetaksikariv]; *tu arrives* [tariv]; TRANEL 2000, p. 59).

mélodie flottante de se réaliser : AIF (*Avoid Integrating Floaters*). Aucune contrainte de fidélité qui militerait pour la réalisation des mélodies flottantes ne se trouve au-dessus d'AIF, mais une contrainte NUCLEUS/V s'assure que le noyau d'une voyelle est bien occupé par une voyelle. Ainsi, la hiérarchie en (14.19),⁶ signifie qu'il est pire qu'une syllabe n'ait pas de voyelle dans son noyau que d'insérer une structure permettant la réalisation d'une mélodie flottante. Cette hiérarchie est responsable de l'élision devant initiale vocalique (NUCLEUS/V est satisfait par la voyelle initiale du second mot), et de l'absence d'élision devant initiale consonantique (NUCLEUS/V doit être satisfait, même si cela implique de violer AIF).

(14.19) NUCLEUS/V \gg AIF (TRANEL 1996)

Tranel génère donc l'élision de deux manières différentes dans ses deux articles. Dans l'article de 2000, une contrainte explicitement anti-hiatique l'emporte sur la réalisation du schwa (14.18), tandis que dans l'analyse de 1996, il n'est pas fait état du hiatus mais du fait qu'une syllabe doit avoir une voyelle pour noyau, ce qui est plus important que la pénalité relative à la réalisation d'une mélodie flottante (14.19). Dans les deux cas, c'est toujours le caractère adjacent de deux voyelles qui génère l'élision, mais la mécanique est différente. En (14.18), la grammaire 'veut' exprimer le schwa mais pas au point de tolérer un hiatus, tandis qu'en (14.19), la grammaire 'veut' ne pas exprimer le schwa, mais pas au point de tolérer un noyau sans voyelle.

L'analyse que nous proposons se trouve en quelque sorte à mi-chemin entre les deux analyses proposées par Tranel : nous postulons, avec TRANEL (2000) (14.18) que la phonologie a besoin d'exprimer le schwa dans certaines conditions, mais entendons démontrer, avec TRANEL (1996) qu'il n'est pas nécessaire de faire appel à un mécanisme de résolution de hiatus. Notre analyse diffère de celles proposées par Tranel pour deux raisons principales, que nous développons dans les sections suivantes. D'une part, l'a.fr. élide d'autres voyelles que schwa beaucoup plus souvent que ne le fait le fr.mod., et, en période 1, nous avons besoin de désigner les voyelles élidables par le flottement hérité de COND CV ; une contrainte ciblant schwa et non une mélodie flottante n'est donc pas satisfaisante pour l'a.fr. D'autre part, en période 1, l'élision a lieu pour les monosyllabes fonctionnels de structure CV alors que les hiatus internes n'ont pas encore commencé leur réduction ; une contrainte générale anti-hiatique n'est donc pas adaptée. Dans les sections suivantes, nous envisageons ce que permet chacune des deux analyses de Tranel, avant de mettre au point notre propre analyse.

14.2.3.2 Flottement et élision en fr.mod. (1) : \emptyset sans V

Si l'on s'en tient aux gabarits proposés en (14.12) *supra*, le fonctionnement de l'élision en a.fr. semble découler naturellement de l'analyse proposée par TRANEL (1996) : la rencontre d'un noyau vide et d'une autre voyelle n'a pas de raison d'être réalisée comme un hiatus puisqu'il faudrait pour cela insérer une position V supplémentaire. Ainsi, en (14.20), sur la base d'une représentation où la mélodie flottante n'est pas alignée avec une position dans le lexique, l'élision est la situation par défaut : aucune

6: Nous simplifions une nouvelle fois la hiérarchie en ne retenant que les éléments pertinents pour notre discussion.

position n'est insérée (AIF chez TRANEL 1996), et la mélodie flottante n'est donc pas réalisée.

(14.20) *le + or* → *l'or*⁷

C	V	C
l	ə	o r

C'est devant initiale consonantique que la solution est moins évidente, puisqu'il s'agit alors d'insérer une position absente de la représentation lexicale du premier mot. Ainsi, dans la représentation d'entrée en (14.21), aucune position n'est disponible pour accueillir la mélodie flottante.

(14.21) *le bastun* → **l bastun*⁸

C	C	V
l	ə	b a ...stun

L'insertion d'une position pour accueillir la mélodie flottante est donc générée par le fait que des contraintes hiérarchisées au-dessus de AIF imposent insertion. Il s'agit pour l'auteur des contraintes PARSE (*Avoid deleting segment*) et FILL (*Avoid inserting segments*). De fait, la hiérarchie complète pour l'élimination est celle donnée en (14.22), que l'on peut gloser ainsi : une syllabe doit avoir une voyelle pour noyau (NUCLEUS/V), et il est pire d'insérer un segment (non flottant ; PARSE) ou de supprimer un segment (id., FILL) que d'insérer une position permettant la réalisation d'une mélodie flottante (AIF).

(14.22) NUCLEUS/V ≫ { PARSE, FILL } ≫ AIF (TRANEL 1996)

La solution optimale pour doter la syllabe d'un noyau est donc l'insertion d'une position permettant à la mélodie flottante de se réaliser. On peut déduire de la hiérarchie en (14.22) la représentation gabaritique en (14.23) ; sur ce schéma, la position V grisée est celle qui résulte de l'insertion.

(14.23) *le bastun* → *le bastun*

C	V	C	V
l	ə	b	a ...stun

Cette solution, qui est celle que l'on peut détailler à partir de l'interprétation de l'article de TRANEL (1996),⁹ est fonctionnelle. Devant initiale vocalique, rien ne se produit, et devant initiale consonantique, une position est insérée pour servir de support à la mélodie flottante.

Néanmoins, il nous semble qu'une alternative est préférable à l'insertion pure et simple d'une position devant initiale consonantique, et ce pour deux raisons. La première raison est que la majorité des mots (occurrences) de l'a.fr. est à initiale consonantique et non vocalique, comme nous avons pu le constater à de multiples reprises en comparant le nombre

7: Par exemple *Roland* v. 185 : *Quatre cenx muls chargez de l'or d'Arabe*, « quatre cents mulets chargés d'or d'Arabie » (trad. SHORR 1990).

8: Par exemple *Roland* v. 341 : *puis li livrat le bastun et le bref*, « [il] lui a remis le bâton et la lettre » (trad. SHORR 1990).

L'alignement vertical entre l'origine de la position V insérée et la position V de *bastun* est fortuite ; sur le schéma, V doit bien venir de quelque part, mais c'est seulement un artifice graphique.

9: Tranel lui-même ne donne aucune représentation gabaritique dans son article ; nos schémas sont issus de l'interprétation du fonctionnement de ses contraintes.

d'occurrences par contexte dans la partie précédente. De ce fait, sa solution reviendrait, pour l'a.fr., à ce que la situation par défaut soit celle dont la fréquence est la moins élevée. Si cette situation n'est pas impossible à admettre dans l'absolu, elle affaiblit néanmoins l'analyse, notamment dans une perspective acquisitionnelle. La deuxième raison, d'ordre purement théorique, est qu'il nous semble préférable d'éviter d'insérer une position à des fins purement phonotactiques, et ce d'autant que dans un des courants théoriques qui a émergé de la phonologie autosegmentale, la phonologie de gouvernement, ce sont précisément les positions qui déterminent les alternances et non l'inverse.

14.2.3.3 Flottement et élision en fr.mod. (2) : ə ⇔ V

La seconde analyse de TRANEL (2000), quoique ne reposant pas sur des voyelles flottantes, répond aux deux objections soulevées ci-dessus. Selon cette analyse, ce n'est plus une position qui est insérée devant initiale consonantique, mais un schwa qui est supprimé devant initiale vocalique. Nous reproduisons, pour rappel, la hiérarchie des trois contraintes pertinentes pour notre analyse :

$$(14.18) \text{ MAX}(V) \gg *VV \gg \text{MAX}(\text{schwa}) \quad (\text{TRANEL 2000})$$

Les représentations gabaritiques que nous pouvons dériver de cette analyse ne contiennent donc pas de schwa flottant ; si l'auteur envisage cette possibilité en note (p. 95, n. 19), il la refuse néanmoins de manière à autoriser des cas où schwa est épenthétique (*extrinsèque*), comme certaines consonnes (le *-t* de *mange-t-il* par exemple). Si l'on accepte provisoirement que les schwas, du moins ceux du fr.mod., ne soient pas flottants, la situation par défaut est donc celle devant initiale consonantique : **VV* n'est pas violée, et donc *MAX(schwa)* n'a aucune raison de l'être. Aussi, en (14.24), rien ne se produit : schwa est lexicalement associé à sa position et rien ne s'oppose au maintien de cette association.

$$(14.24) \begin{array}{cc} \text{C} & \text{V} & & \text{C} & \text{V} \\ | & | & & | & | \\ \text{l} & \text{ə} & & \text{b} & \text{a} \quad \dots\text{stun} \end{array}$$

C'est en revanche devant initiale vocalique que la computation produit un changement vis-à-vis des représentations lexicales. Sans ce changement, *le + or* provoquerait *le or* (14.26), ce qui n'est pas la configuration attestée dans notre corpus. Au contraire, l'élision de l'article *le* est systématique (99%). C'est ici que la hiérarchie **VV* \gg *MAX(schwa)* produit son effet : le hiatus est moins tolérable que la suppression d'un schwa, ce qui entraîne, en (14.26), la suppression du schwa.

$$(14.25) \text{ le + or} \rightarrow *le \text{ or} \quad (14.26) \text{ le + or} \rightarrow l'or$$

C	V	V	C
l	ə	o	r

C	V	V	C
l	ə	o	r

14.2.3.4 Mécanisme retenu pour l'a.fr. : ə ⇔ V

Comme indiqué *supra*, la variété des voyelles qui s'élident est plus grande en a.fr. qu'en fr.mod. : /ə/ dans *me, te, le*, /a/ dans *la*, /o/ dans *jo*, /i/ dans *li* sg. Aussi, disposer d'un mécanisme qui détruirait spécifiquement les schwas en hiatus n'est pas adapté à la situation de l'a.fr., où l'on ne peut admettre, comme le fait Tranel pour le fr.mod., un supplétisme des formes en schwa devant initiale vocalique. Et ce d'autant que les autres supplétismes qui corroborent le remplacement de *la* par *le* devant initiale vocalique pour Tranel sont absents en a.fr. En effet, le remplacement du féminin par le masculin pour, par exemple, les possessifs, n'est pas encore actif en a.fr., et les possessifs féminins s'élident naturellement (*ma espee* → *m'espee*¹⁰ en a.fr. devient *mon épée* en fr.mod., *ma amie* → *m'amie*¹¹ devient *mon amie*; id. *s'espee*,¹² *s'amie*,¹³ *t'amie*¹⁴, etc.). Par ailleurs, nous avons besoin de pouvoir encoder la distinction entre d'une part *li* pl et *tu* CS et d'autre part tous les autres monosyllabes fonctionnels, et nous avons avancé en sec. 14.2.1.1 que le flottement de la mélodie était un outil idéal pour ce faire. Nous retenons donc, dans notre analyse, le fait que les voyelles finales sujettes à élision sont flottantes.

Néanmoins, nous cherchons également à éviter l'insertion d'une position gabaritique qui permettrait de réaliser la voyelle flottante devant initiale vocalique, de manière à respecter le fait que la situation par défaut est, statistiquement, la non-élision devant initiale vocalique. Une solution qui permet de résoudre ces deux problèmes est d'emprunter à certains développements de la phonologie de gouvernement l'idée que la structure syllabique est toujours rapportée à une stricte alternance de positions C et V (LOWENSTAMM 1996). Si l'on accepte ce principe, il n'est plus possible de choisir parmi les différentes représentations du flottement exposées en (14.5) *supra* : dans le cas de nos monosyllabes fonctionnels, la mélodie flottante flotte nécessairement *sous* une position V, tel qu'en (14.27), qui reformule les représentations en (14.9).

(14.27) Voyelles flottantes en synchronie

(a)	C	V	(b)	C	V	(c)	C	V	(d)	C	V
	l	ə		l	i		j	o		t	u

Dans ce cas, devant initiale consonantique, la mélodie flottante dispose d'une position disponible à laquelle s'associer, et la seule chose qui manque est le mécanisme d'association. Nous proposons une version simple de ce mécanisme, intitulée AssocV et définie selon (14.28). Ce mécanisme est en apparence exactement inverse à l'AIF utilisé par TRANEL (1996) : il promeut la réalisation d'une mélodie flottante par l'association de celle-ci avec une position disponible. Mais cette contradiction entre AssocV et AIF n'est que partielle : ce que prohibe AIF, ce n'est pas la réalisation d'une mélodie flottante, mais bien l'insertion d'une position qui lui permettrait d'être réalisée. Or, ici, la position est déjà disponible, et AssocV ne fait qu'établir la ligne d'association entre mélodie et position. Cette ligne d'association, absente du lexique et due à la computation, est représentée par une flèche ↑ en (14.29).

10: P.ex. *Roland*, v. 620, 966, 1901.

11: P.ex. *Perceval*, v. 2267, 3865, 4032, 4431, 8301, 8497, 8523, 8531, 8545, 8855.

12: P.ex. *Roland*, v. 346, 465, 1570, 1629, 2089, 2259, 2264, 2283, 2572.

13: P.ex. *Perceval*, v. 784, 2573, 2623, 2910, 2931, 3980, 4179, 4188, 4338, 4670, 4968, 4999, 6897, 8745, 8775.

14: P.ex. *Perceval*, v. 3921

LOWENSTAMM (1996). « CV as the only syllable type ».

(14.28) AssocV

Lorsqu'une mélodie vocalique flottante peut, sans croiser de ligne d'association, s'associer avec une position V vide, elle le fait.

(14.29) *le bastun* → *le bastun*

C ₁	V ₁	C ₂	V ₂	
	↑			
l	ə	b	a	...stun

C'est en revanche devant initiale vocalique qu'un problème se pose. En (14.30), si l'on admet la position V₁ au-dessus de la mélodie flottante, rien n'empêche AssocV d'associer la mélodie à la position, ce qui crée alors un hiatus. Plutôt que d'appliquer d'autorité un mécanisme anti-hiatique qui interdirait la succession de deux mélodies vocaliques et/ou la succession de deux positions V associées à une mélodie, nous pensons que la solution réside dans le fait d'accepter que l'organisation des positions fonctionne, dès le lexique, selon une stricte itération de couples de positions CV. De ce principe, qui voit en CV le type syllabique universel (LOWENSTAMM 1996), découle la conclusion que les mots à initiale vocalique commencent par une position C vide, telle que C₂ en (14.30).

(14.30) *le + or* → **le or*

C ₁	V ₁	V ₂	C ₂
	↑		
l	ə	o	r

(14.31) *le + or* → **le or*

C ₁	V ₁	C ₂	V ₂	C ₃	V ₃
	↑				
l	ə	o	r		

L'intégration de C₂ à la représentation (14.30) ne résout pas immédiatement la situation; AssocV peut toujours associer la mélodie flottante à la position V₁. Or en (14.31), on trouve la succession d'une position V vide et d'une position C vide, configuration que nous désignons au singulier comme un VC vide. Dans le cadre de la phonologie de gouvernement, GUSSMAN et KAYE (1993) ont proposé un mécanisme universel¹⁵ de suppression des ces configurations VC vides. Le mécanisme, appelé *reduction*, est défini comme tel :

Reduction

An empty nucleus followed by a pointless onset are removed from any phonological representation in which they occur.

(GUSSMAN et KAYE 1993, p. 433)

Ce mécanisme, que nous abrégons *VCvides, et que nous reformulons tel qu'en (14.32), est au cœur de l'analyse que nous proposons : il permet de dériver l'élision dans tous les contextes où celle-ci est attestée, et de ne pas la dériver dans les contextes où elle n'est pas attestée.

(14.32) *VCVIDE

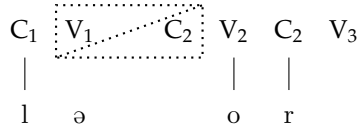
Toute séquence de positions VC où V et C sont vides est éliminée.

GUSSMAN et al. (1993). « Polish notes from a Dubrovnik Café: I. The yers. »

15: GUSSMAN et KAYE (1993, p. 433) : « We assume that it is a part of UG and not a peculiarity of Polish. »

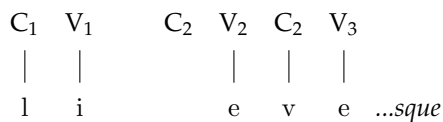
Comme illustré en (14.33), l'application de *VCVIDES prive la mélodie flottante de toute position V à laquelle elle pourrait s'attacher; AssocV (14.28) n'autorise pas l'association de la mélodie flottante en V₂, qui n'est pas vide.

(14.33) *le + or → l'or*



La réduction des VC vides, qui est devenue relativement commune en phonologie de gouvernement (et en CVCV, l'un des types de phonologie de gouvernement),¹⁶ se révèle donc capable de générer l'élision devant initiale vocalique. Ce mécanisme, contrairement à une contrainte générale du type *VV, ne cible aucun hiatus qui ne contiendrait pas de mélodie flottante. Il parvient ainsi à respecter la valeur diacritique de l'association entre position et mélodie pour *li pl* et *tu CS*. Ainsi, si l'élision procède de l'élimination d'un VC vide, elle ne peut pas s'appliquer en (14.34), où V₁ n'est pas vide.

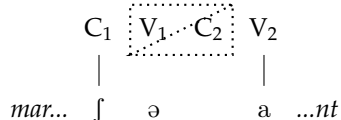
(14.34) *li + evesque → li evesque*¹⁷



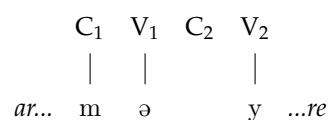
Externalité positive de *VCVIDES (1) : hiatus internes. Comme mentionné en sec. 14.2.2.1, nous proposons la règle synchronique COND.ə (14.17), qui rend flottants tous les schwas. *VCVIDES permet également de générer la réduction de ces hiatus, et permet de prendre en compte la lente diffusion du phénomène dans le lexique. Il suffit pour cela de postuler que COND.ə met du temps avant d'atteindre tous les lemmes hiatiques. Tant que le schwa hiatique n'est pas flottant, *VCVIDES n'a pas de cible (la position V est associée); dès que le schwa est flottant, la règle peut s'appliquer.

Ainsi, dans *nimd*, le lemme « marchéant » est déjà atteint par la réduction en *marchant* (p. 27, 35), tandis que ce n'est pas encore le cas du lemme « armëure » (p. 4). Il suffit donc d'appliquer COND.ə au lemme « marchéant » (14.35) mais pas encore au lemme « armëure » (14.36), pour obtenir, dans le même état de la langue, la réduction du hiatus par *VCVIDES dans l'un et pas dans l'autre.

(14.35) *marchéant → marchant*



(14.36) *armëure → armëure*



16: Cf. p.ex. FAUST, LAMPITELLI et ULFSBJORNINN 2018, p. 11; FAUST 2023a, p. 127; FAUST et LAMPITELLI 2017, p. 49; FAUST 2023c, p. 6, 2023b, p. 22.

17: Par exemple *hyla* p. 28 : *li evesque li prestre et li moine venoient a lui de toutes parties*, « les évêques, les prêtres et les moines venaient à lui de toute part » (trad. pers.). **hyla** : WAUCHIER DE DENAIN (1208*). *La vie de saint Hylaire*. Dees : Somme, Pas-de-Calais ; compo. Belgique, ca 1208 ; ms. pic., ca 1275 ; MS1 (transcr. inédite Th. Venckeleer).

nimd : ANON. (1150*). *Le charroi de Nîmes*. Ms. D ; Dees : Meuse ; compo. nil, ca 1150 ; ms. lorr. mérid., ca 1275 ; MS1 (transcr. inédite L. Schoesler).

***VCVIDES n'est pas un mécanisme anti-hiatique.** Dans notre utilisation de *VCVIDE, la règle a un effet anti-hiatique. C'est également le rôle que lui assigne FAUST (2023b, p. 22) : « Delete problematic VC sequences (i.e. those that causes hiatuses). » Mais ce mécanisme, du moins dans sa formulation originelle par GUSSMAN et KAYE (1993), ne saurait se réduire à cela : les auteurs l'utilisent pour obtenir une structure qui prohibe une relation de gouvernement propre. Dans leur analyse des *yers*, voyelles tchèques alternant avec zéro, les auteurs considèrent avec SPENCER (1986) que les *yers* sont des voyelles épenthétiques, représentées par un simple noyau vide. Les *yers* apparaissent lorsqu'une position consonantique N n'est pas finale, n'est pas la cible d'un gouvernement propre et ne précède pas un groupe *s+C*; une position licenciée ne peut pas émettre de gouvernement propre. En (14.37 - 14.39), une position N est une position finale (de cycle) et l'épenthèse de *yer* n'a pas lieu.

La structure en (14.37) représente l'étape [[pØsØ] ØkØ] de la dérivation de *pes* (« chien ») en *pieseczek* (double diminutif).¹⁸ Le but de la réduction, ici, est de pratiquer l'épenthèse de *yer* sur la deuxième position N, pour obtenir à l'étape suivante la forme *pieseczek* et non la forme *piesØczek*. Or, en (14.37), N₄ peut gouverner N₅, ce qui dispense N₅ d'être sujet à l'épenthèse. C'est donc en appliquant la réduction (qui élimine N₅ O) que Gussman et Kaye obtiennent la structure en (14.38), où la deuxième position N, N₄, ne remplit pas les conditions lui permettant d'échapper à la réalisation du *yer*. Nous représentons en (14.39) le passage de (14.37) à (14.38), en encadrant et barrant les positions supprimées par la réduction.

18: Nous ne justifions pas la structure des morphèmes ni cycles postulés par les auteurs.

(14.37) Structure avant réduction

O	N ₆	O	<u>N₅</u>	O	N ₄	O	N ₃
×	×	×	×	×	×	×	×
p		s				k	

(GUSSMAN et KAYE 1993, p. 433)

(14.38) Structure après réduction

O	N ₆	O	N ₄	O	<u>N₃</u>
×	×	×	×	×	×
p		s			k

(GUSSMAN et KAYE 1993, p. 434)

(14.39) Structure montrant les positions éliminées

O	N ₆	O	<u>N₅</u>	Ø	N ₄	O	<u>N₃</u>
×	×	×	×	×	×	×	×
p		s				k	

Le processus que nous interprétons comme *VCVIDE ne saurait donc se réduire, du moins dans l'analyse de GUSSMAN et KAYE (1993), à un simple procédé anti-hiatique; il trouve son origine dans le besoin de gérer correctement des relations latérales. La dimension anti-hiatique

de *VC_{VIDE} n'est qu'une conséquence du processus, et ne saurait en être le seul moteur. *VC_{VIDE} n'est donc pas équivalent au *VV de TRANEL (2000). Chez Tranel, *VV cible toutes les voyelles, et ce que par l'effet de MAX(V) \gg *VV \gg MAX(schwa) que *VV n'atteint pas les autres voyelles que schwa. L'analyse que nous proposons avec CONDCV, COND.ə, ASSOCV et *VC_{VIDE} ne procède donc pas d'une réduction générale des hiatus; elle ne génère que l'effacement des mélodies vocaliques flottantes devant voyelle (initiale ou interne), par l'effet d'un processus indépendamment motivé.

14.2.3.5 Note sur *que*

La conjonction *que* fait exception aux données avancées ci-dessus : son taux d'élision est inférieur à celui des articles, pronoms et prépositions *le*, *la*, *te*, *me* et *de*, avec 69% d'élision pour le *-que* des locutions conjonctives en *base* + *-que* et 41% pour celui de l'élément *que* autonome. Si, comme expliqué *supra*, ce taux de 41% pour l'élément *que* autonome est biaisé par la présence de pronoms conjonctifs (y compris *qui*) dans les données, le taux de 69% pour les locutions est en revanche plutôt fiable. De fait, la non élision de *que* n'est pas rare, et est même l'un des cas de déni d'élision les plus saillants à la lecture de textes versifiés.

Cette singularité de *que* peut s'expliquer si l'on remarque qu'en t.a.fr., *que* est pourvu d'une consonne de liaison (QUOD \rightarrow /k^{od}/). Dans *Eulalie*, cette consonne n'est pas systématique : on trouve deux occurrences *quelle* pour une occurrence *qued elle* (v. 14). Dans les parties romanes du *Sermon sur Jonas*, la forme graphique élidée est absente, et l'on trouve soit *que* (abrégé *q̄*) soit *quet* (abrégé *q̄t̄*) devant initiale vocalique. Dans *Alexis*, l'opposition n'est pas systématique non plus : on trouve *qu'il* (v. 340) et *qued id* (v. 103).

Ces attestations s'accompagnent d'autres alternances consonantiques et vocaliques, toujours en fin de mot fonctionnel monosyllabique, qui sont à analyser comme des cas de liaison consonantique, comme en (14.40 a - 14.40 b), et de cas d'élision (14.40 c - 14.40 d, et d'autres, cf. CHASLE 2008; PREMAT 2022). Si, dans *Eulalie*, l'élision n'est attestée que pour *que*, elle est attestée à deux occurrences dans *Strasbourg*,¹⁹ et est systématique pour *le*, *la*, *te*, etc. dès *Alexis*.

Elles démontrent donc deux choses : d'une part les phénomènes de sandhi externe entre clitiques (monosyllabes fonctionnels de structure archaïque CV) et mot lexical sont actifs dès les premiers textes (*d* et *l* dans *Strasbourg*). D'autre part, la conjonction *que*, qui n'atteint pas 100% d'élision dans notre corpus d'a.fr., descend d'une conjonction polymorphe, pouvant avoir les formes *que*, *quet* ou *qued*, et *qu*. Dans la lignée des travaux de CLEMENTS et S. KEYSER (1983), ENCREVÉ (1983, 1988) et TRANEL *passim*, nous postulons pour la consonne de liaison qui apparaît dans *qued* et *quet* le même statut de flottement que pour les voyelles finales des clitiques, soit la représentation en (14.41).

(14.41) *qu(e)(d)* : représentation lexicale (1)

C	V	C	V
k	ə	d	

eulaLiBfm : ANON. (881*). *Séquence de sainte Eulalie*. Compo. peu après 881, ms. ca 900, (LAVRENTIEV, GUILLOT-BARBANCE et RAINSFORD 2018).

jonas : ANON. (ca 945*). *Sermon sur Jonas*. Compo. prob. entre 938 et 952, ms. 2^e q. x^e s., dial. fl. (POERCK 1956).

CHASLE (2008). « Manifestation de la latence en ancien français aux x^e et xi^e siècles : liaison et redoublement syntaxique ».

PREMAT (2022). « Le sandhi externe en très ancien français ».

19: *De ist* \rightarrow *d'ist*, « depuis ce [jour] », l. 5; *le int* \rightarrow *l'int*, « l'en [détourner] », l. 31; trad. pers.

jonas : ANON. (ca 945*). *Sermon sur Jonas*. Compo. prob. entre 938 et 952, ms. 2^e q. x^e s., dial. fl. (POERCK 1956).

eulaLiBfm : ANON. (881*). *Séquence de sainte Eulalie*. Compo. peu après 881, ms. ca 900, (LAVRENTIEV, GUILLOT-BARBANCE et RAINSFORD 2018).

AlexisRaM : ANON. (1050*). *Vie de saint Alexis*. Compo. ca 1050, ms. ca 1120, dial. norm. (RAINSFORD et MARCHELLO-NIZIA 2018).

roland : ANON. (ca 1100). *Chanson de Roland*. Compo. ca 1100, ms. ca 1120, dial. norm. (MOIGNET 1972a).

Cette représentation se doit d'avoir, en synchronie du t.a.fr., à la fois /ə/ et /d/ flottants, puisque dès que *qued* est attesté dans *Eulalie*, qu'il est aussi. Mais une telle structure échouerait à devenir clitique au vu de COND_{CV} : /kəd/, ou /kə^d/ (l'exposant indique le flottement) tel qu'en (14.41), correspondent à un gabarit CVCV pour une forme CVC. Si l'on veut maintenir la possibilité que COND_{CV} déclare le schwa de *que(d)* flottant, il faut envisager que c'est la représentation en (14.42) qu'évalue COND_{CV} en proto-a.fr., ce qui engendre la représentation lexicale en (14.43).

(14.40) Sandhi externe en t.a.fr.

	Lat.		__#C	__#V		
(a)	« Et »	ET	→	<i>e</i>	<i>et</i> (abrégé $\mathcal{T} = \&$)	<i>Jon., Alex.</i>
(b)	« À »	AD	→	<i>a</i>	<i>ad</i>	<i>Eul., Jon., Alex., Rol.</i>
(c)	« De »	DE	→	<i>de</i>	<i>d</i>	<i>Stras., Alex., Rol.</i>
(d)	« Le, la »	ILLE, -A	↔	<i>le, la</i>	<i>l</i>	<i>Stras., Alex., Rol.</i>

(CHASLE 2008; PREMAT 2022)

Ainsi, si l'on admet que c'est la représentation lexicale en (14.42) qui a

(14.42) *que(d)* : représentation (2) avant COND_{CV} (14.43) *qu(e)(d)* : représentation après COND_{CV}

C	V		C	V	
k	ə	d	k	ə	d

été testée par COND_{CV} en proto-a.fr., le mot dispose bien d'une voyelle flottante, et rien n'impose, en soi, que sa consonne flottante soit linéarisée sous une position C au sein du même mot (14.43). En très ancien français, les deux possibilités *qued* et *qu* semblent possibles devant initiale vocalique. Si l'on admet que le gabarit lexical auquel est associé le matériel mélodique /kə^d/ n'a la forme que d'un seul CV, on obtient donc la situation suivante devant initiale vocalique (14.44).

(14.44) *qu(e)(d) elle* : représentation *a priori*

C	V		C	V	C	V
					↑	
k	ə	d	e	l	ə	

Or en (14.44), il y a bien un VC vide. Diachroniquement, la disparition nette des formes du type *qued* ou *quet* après *Alexis* montre que la computation semble avoir privilégié *VC_{VIDES}. Mais, en très ancien français, la présence des deux types permet soit l'élision, déclenchée par *VC_{VIDES} (14.45), soit la forme avec consonne de liaison, contraire à *VC_{VIDE} (14.46).

Face aux attestations *qued* du très ancien français, nous sommes contraints d'accepter que *VC_{VIDES} ne s'applique pas systématiquement. La clef, ici, tient peut-être à la présence de deux mélodies flottantes. Elle peut aussi tenir au fait que *VC_{VIDES} n'en est peut-être qu'au début de sa période d'activité, et n'a pas encore atteint le stade de la systématité qu'elle aura en a.fr. Le cumul des deux facteurs (deux mélodies vides et potentielle défaut de jeunesse de *VC_{VIDES}) peut être conservé comme des éléments d'explication : en effet, si les formes *qued* et *quet* disparaissent

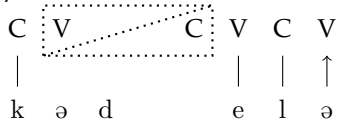
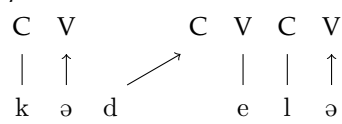
jonas : ANON. (ca 945*). *Sermon sur Jonas*. Compo. prob. entre 938 et 952, ms. 2^e q. x^e s., dial. fl. (POERCK 1956).

AlexisRaM : ANON. (1050*). *Vie de saint Alexis*. Compo. ca 1050, ms. ca 1120, dial. norm. (RAINSFORD et MARCHELLO-NIZIA 2018).

eulaLiBfm : ANON (881*). *Séquence de sainte Eulalie*. Compo. peu après 881, ms. ca 900, (LAVRENTIEV, GUILLOT-BARBANCE et RAINSFORD 2018).

roland : ANON. (ca 1100). *Chanson de Roland*. Compo. ca 1100, ms. ca 1120, dial. norm. (MOIGNET 1972a).

strasBfm : ANON (842*). *Serments de Strasbourg*. Compo. ca 842, ms. ca 1000, (LAVRENTIEV, RAINSFORD et GUILLOT-BARBANCE 2014).

(14.45) *qu elle*(14.46) *qued elle*

entre le t.a.fr. et l'a.fr., il n'en demeure pas moins que l'élision de *que* est moins systématique que celle des autres monosyllabes à voyelle finale flottante.

Ainsi, si nous ne sommes pas en mesure de proposer une explication formelle convaincante pour les refus d'élision de *que*, plus nombreux que ceux des autres clitiques, nous avons le sentiment que la présence d'une ancienne consonne de liaison ou, éventuellement, d'un gabarit potentiellement CVCV (14.41) concurrent du gabarit C^V (14.46) pourraient expliquer que les locuteurs et scripteurs aient une conscience, ne serait-ce que partielle, du fait que *qued* ne s'élide pas toujours voire n'est pas toujours traité comme un clitique. Les cas de non élision de *que* seraient alors comme des réminiscences de l'ancienne option *qued*.

Enfin, ces formes non élidées de *que* sont présentes sur une aire géographique définie, le long d'une diagonale partant de l'Angleterre au Nord-Ouest jusqu'à la Bourgogne au Sud-Est, comme représenté en fig. 14.2.²⁰ Les territoires en marge de cette diagonale ne connaissent que très peu le déni d'élision de *que*, ce qui impliquerait que la grammaire de ces variétés a régularisé l'application stricte de *VCVIDES (14.45) au détriment d'une réminiscence de *qued*. Une partie du domaine d'oïl n'est donc pas véritablement concernée par cette variation *qu* ~ *que* devant initiale vocalique, la grammaire ayant évacué la tension qui, ailleurs, génère encore des formes *que* devant initiale vocalique. De manière assez surprenante, cette distribution en deux zones discontinues indique que la non élision de *que* est probablement une innovation et non un trait conservateur. Les périphéries où *que* résiste moins à l'élision représenteraient alors des aires reliquaires (CHAMBERS et TRUGDILL 1998, p. 94), devenues discontinues du fait de l'extension d'une innovation — la possibilité que *que* refuse l'élision. Il est donc possible que les formes *que* devant initiale vocalique ne soient pas le 'descendant naturel' des alternances *qued* ~ *qu* du t.a.fr., mais soient en quelques sortes un archaïsme, qui n'est pas autochtone partout, et dont l'expansion suit un axe Nord-Ouest - Sud-Est.

20: Cf. sec. 10.2.2 pour le commentaire détaillé de la distribution diatopique.

CHAMBERS et al. (1998). *Dialectology*.

14.2.4 En somme

Dans cette section, nous avons proposé de transcrire le diacritique qui distingue, en période 1, les monosyllabes fonctionnels qui s'élident de ceux qui ne s'élident pas, par le flottement de leur mélodie vocalique finale ; c'est ce flottement que nous avons appelé *flottement diacritique* (14.2). Nous avons avancé que c'était ce flottement de la mélodie vocalique finale qui conditionnait l'élision, en adoptant provisoirement ÉLID.FLOT, une condition qui prévoit que seules les voyelles flottantes sont candidates à

l'élision (sec. 14.2.1.1), puis avons reformulé la règle COND_{CV} qui s'est appliquée en proto-a.fr. et a déterminé le flottement des voyelles finales atones des monosyllabes (14.2.1.2). Nous avons ensuite introduit un deuxième type de flottement des mélodies vocaliques, encore actif en synchronie en a.fr., et dénuée de toute dimension diacritique. Il répond à la règle COND_ə, qui rend flottants les schwas et uniquement les schwas (sec. 14.2.2), mais pas uniquement en position finale, dans la mesure où il peut être responsable de la réduction des hiatus internes également (14.2.2.1).

Ayant défini les conditions phonologiques de l'élision, nous avons proposé une analyse de la façon dont le phénomène se déroule (sec. 14.2.3). Nous sommes partis de deux analyses proposées par Tranel, et, après avoir discuté des différences de ces deux analyses (sec. 14.2.3.1 - 3), nous avons repris à notre compte l'idée que c'est la voyelle flottante qui déclenche l'élision, mais que celle-ci ne résulte pas nécessairement d'un processus anti-hiatique. En intégrant certains postulats de la phonologie de gouvernement, nous avons développé une analyse du phénomène où tout mot commence et se termine par des séquences CV, où toute mélodie vocalique flottante est associée à une position V vide si une telle position est disponible (AssocV), et où les VC vides sont supprimés (*VC_{VIDES}). Ces deux mécanismes, si l'on accepte les représentations lexicales utilisées, génèrent efficacement les élisions attestées dans notre corpus.

La vision d'ensemble de l'élision qui se présente à ce stade est la suivante. Deux processus phonologiques ont rendu flottantes certaines mélodies vocaliques finales : COND_{CV}, qui s'est appliqué en proto-a.fr. et a rendu flottantes les voyelles finales de monosyllabes atones de structure CV, et COND_ə, qui est active en a.fr. et qui rend flottant tout schwa. Devant initiale consonantique, le schwa ne s'élide pas : une position V est disponible et il y est attaché par AssocV. Devant initiale vocalique, un VC vide est présent, et celui-ci est éliminé avant que AssocV n'ait eu l'occasion d'associer la mélodie flottante, ce qui provoque l'élision.

14.3 Domaines de l'élision

Dans la section précédente, nous avons défini les conditions phonologiques de l'élision et la façon dont le processus s'applique. Il faut maintenant détailler les domaines qui régissent l'élision, et notamment la différence entre le domaine de la période 1 et celui de la période 2. Nous commençons par définir le domaine de la période 1 comme étant celui de la cliticité (sec. 14.3.1), avant d'envisager ce que peut être le domaine dans la période 2 et la transition de l'un à l'autre (sec. 14.3.2).

14.3.1 Période 1 : élision des clitiques

Nous avons établi au chap. 12 que l'élision de la période 1 ne concernait que des mots fonctionnels monosyllabiques. Cette catégorie correspond donc aux traits habituellement associés à la cliticité : être un mot fonctionnel et non un mot lexical, et ne pas être trop grand. Dans cette section, nous proposons donc une définition de ce que la première période de l'a.fr.

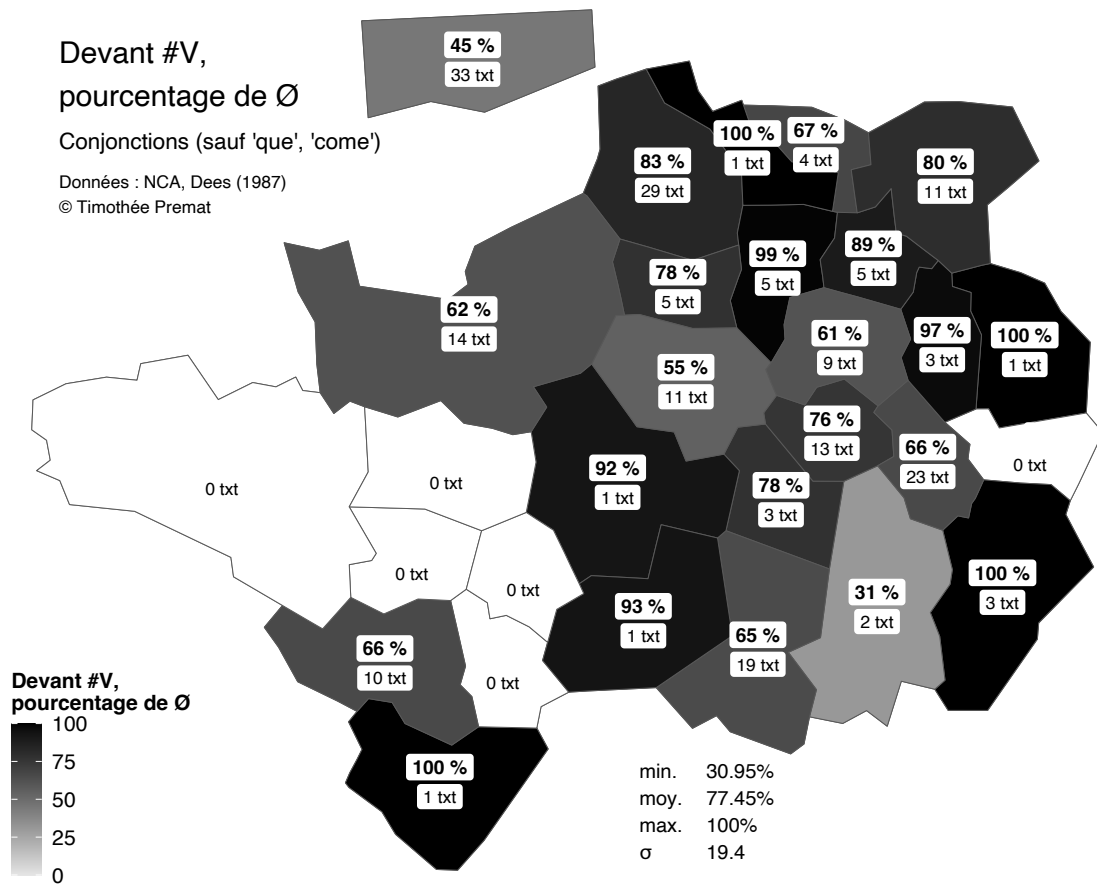


FIGURE 14.2 – Distribution diatopique des alternances <e>/Ø, sur l'ensemble de la période considérée, conjonctions en *base* + *-que* suivies d'une initiale vocalique.

traite comme un clitique. Nous ne traitons pas, ici, des processus d'enclise. Si les mêmes règles ou des règles analogues peuvent être à l'œuvre pour l'enclise et pour la proclise, l'enclise demeure dans le hors champ de notre analyse.

La cliticité est donc liée à la notion de mots fonctionnels. La distinction entre mots fonctionnels (en classe fermée) et mots lexicaux (en classe ouverte) est *a priori* présente dans toute langue, et cette distinction préside à celle entre clitiques et non-clitiques. En effet, alors que les mots lexicaux constituent la situation par défaut, les mots fonctionnels sont susceptibles d'être traités différemment par la phonologie, *i.e.* de ne pas disposer d'une autonomie phonologique et d'être clitiques (SELKIRK 1996, p. 187). Dans le cadre de la phonologie prosodique classique,²¹ le traitement des clitiques est gérée en négatif : « *the set of constraints governing the interface between morphosyntactic and prosodic structure makes no reference to functional categories at all* » (*id.*, p. 191, italiques de l'auteur).²² En conséquence, un clitique ne constitue pas *a priori* un domaine computationnel autonome, et doit être intégré au domaine d'un autre mot. Si notre analyse ne dépend pas d'une analyse en phonologie prosodique, nous reproduisons ce

SELKIRK (1996). « The prosodic structure of function words. »

21: *Phonologie prosodique classique* est une traduction de ce que SELKIRK (2011) appelle *the 'standard theory' of prosodic structures*, correspondant au corpus classique de la phonologie prosodique des années 80 (SELKIRK 1981, 1986; NESPOR et VOGEL 2007 [1986]). Si ce corpus n'est pas sans diversité, il partage néanmoins un certain nombre de fondamentaux communs, comme l'utilisation d'arbres (ou de parenthésages) répondant à la hiérarchie prosodique et l'hypothèse de l'étagement strict (*strict layering*).

principe selon lequel la cliticité résulte d'une absence de considération de certains mots fonctionnels par la structure de domaines phonologiques : les domaines phonologiques ne sont pas établis en fonction de ces items, qui deviennent alors clitiques. Cette idée n'est d'ailleurs pas propre à la phonologie prosodique, et peut être implémentée par toute théorie de l'interface entre morphosyntaxe et phonologie, que ce soit au moyen de strates, de phases, etc. Pour nous, du moins dans le cadre de ce manuscrit, peu importe la nature du domaine ; seule importe la définition de ces cibles.

Admettons que dans la période 1 de l'a.fr., le domaine computationnel ne corresponde qu'à un seul mot lexical ; c'est ce qui est sous-tendu par toutes les descriptions de l'*accent de mot* hérité du latin, de l'absence²³ de liaison consonantique et, dans nos données graphiques, de l'absence²⁴ d'élision des mots fonctionnels polysyllabiques. Tel que formulé en (14.47), le principe de constitution des domaines computationnels ne dit rien des mots fonctionnels ; il ne prédit donc pas ce qu'il advient des nœuds terminaux syntaxiques qui ne sont pas des mots lexicaux.

(14.47) $DOM \Rightarrow MOTLEX$
Un domaine computationnel correspond à un mot lexical.

Trois situations sont envisageables, à partir de nos données, pour traiter l'absence de prise en compte des mots fonctionnels dans la définition des domaines à partir de la structure syntaxique. La première, qui est fictive, serait celle où, du fait de $DOM \Rightarrow MOTLEX$ (14.47), un clitique serait hors de tout domaine computationnel (14.48). L'impossibilité de cette situation résulte simplement du fait que le domaine computationnel est par définition l'espace dans lequel la phonologie interprète les informations qui lui sont données ; si un mot fonctionnel est hors de tout domaine computationnel, il ne peut être interprété et disparaît.

(14.48) $*DOM = MOTFONC [MOTLEX]$
Pour être traité par la phonologie, un mot doit appartenir à un domaine computationnel.

L'absence d'interprétation ne pouvant être envisagée, il faut donc que le mot fonctionnel se trouve dans un espace qui permette son interprétation. Nous postulons que l'espace dans lequel le mot fonctionnel est interprété peut être indépendant ou dépendant du domaine des mots environnants, selon deux types de critères. Le premier tient à la relation syntaxique entre le mot fonctionnel et le mot lexical le plus proche à sa droite. Le second tient à la taille du mot fonctionnel.

14.3.1.1 Condition de proximité syntaxique

Un mot fonctionnel peut appartenir au domaine du mot lexical le plus proche sur sa droite s'il entretient une relation syntaxique suffisamment *intime* avec celui-ci. Nous utilisons à dessin le mot d'*intimité* pour ne pas présupposer une théorie donnée de la syntaxe ; cette notion pré-théorique peut être transcrite en termes de proximité entre le nœud terminal du mot fonctionnel et le nœud commun le plus bas que le mot fonctionnel partage avec le mot lexical le plus proche à sa droite. Autrement dit, ce nœud

22: Cf. SELKIRK 1984, 1986 parmi d'autres.

23: Ou de l'extrême marginalité.

commun ne doit pas être trop élevé dans l'arborescence syntaxique, vis-à-vis du nœud terminal du mot fonctionnel. Lorsque le mot fonctionnel et le mot lexical partagent un nœud commun assez bas dans l'arbre syntaxique, le mot fonctionnel peut être incorporé au domaine du mot lexical (14.49).

(14.49) DOM = [MOTFONC MOTLEX]

Un mot fonctionnel peut appartenir au domaine du mot lexical suivant lorsqu'il a un nœud commun avec celui-ci de niveau suffisamment bas dans l'arbre syntaxique.

p.ex. *suz l ombre d un pin* (Roland, v. 407)

Cette situation est celle que la phonologie prosodique classique décrit sous l'étiquette de *groupe clitique* (GC : NESPOR et VOGEL 2007 [1986], p. 145-163) — nous ne nous attardons pas sur le débat concernant (a) l'existence d'une structure réursive de type [MOTFONC [MOTLEX]], qui viole l'hypothèse de l'étagement strict (*stric layering hypothesis*; un GC ne peut dominer que des mots prosodiques), (b) son remplacement par une structure de type [[MOTFONC] [MOTLEX]] qui fait du clitique un équivalent prosodique du MOTLEX (cf. NESPOR et VOGEL 2007 [1986], p. xv-xix) ou (c) la pertinence de cette catégorie dans l'absolu (BOOIJ 1986). Plus généralement, l'*intimité syntaxique* évoquée *supra* entre le nœud terminal du clitique et celui du mot lexical vers lequel il penche rend compte de la proximité, ou *degré de connexion*,²⁵ des deux termes dans la structure syntaxique, qui préside à l'application des règles de sandhi externe. Ce degré de proximité peut être formalisé d'une multitude de manières, directement dans une hiérarchie prosodique stricte (NESPOR et VOGEL 2007 [1986], p. 145-163) ou réursive (NESPOR et VOGEL 2007 [1986], p. xv-xix; SELKIRK 2011), ou par le biais d'une grille métrique (SELKIRK 1984), de strates ou de phases. En ce qui nous concerne, nous n'avons pas besoin de définir la façon dont cette *intimité syntaxique* est dérivée de la syntaxe et de paramètres arbitraires; il nous suffit de considérer que dans certains cas, un MOTFONC est placé dans un domaine computationnel qui comprend également le mot lexical le plus proche sur sa droite (ou sur sa gauche dans le cas de l'enclise, dont nous ne traitons pas ici), ce qui autorise l'application du sandhi externe. Cette cliticité ne concerne évidemment pas que l'élision; elle concerne également les cas de liaison consonantique du t.a.fr. illustrés en sec. 14.2.3.5.

24: *Idem.*

Dans l'exemple donné en (14.49), *suz l ombre d un pin*, on peut considérer que l'élision est le symptôme de la communauté de domaine computationnel entre « le » et « ombre », et entre « de » et « un pin ». On peut dans un second temps en déduire que toutes les structures DÉT + NOM et PRÉP + DÉT + NOM sont transmises à la phonologie sous la forme d'un seul domaine computationnel — a ceci près qu'une autre restriction s'applique, selon la taille des MOTFONC (sec. suivante). Dans la mesure où les articles et prépositions monosyllabiques dont la voyelle finale est flottante s'élident presque systématiquement, il semble raisonnable de postuler qu'articles et prépositions sont toujours suffisamment intimes avec le mot lexical qui les suit pour pouvoir appartenir à son domaine.

La dernière possibilité consiste à placer le mot fonctionnel au sein de son propre domaine computationnel. Nous l'envisageons comme une possibilité de dernier recours, activée lorsqu'il n'est pas possible de placer le mot fonctionnel dans le domaine du mot lexical le plus proche à sa

droite. C'est par cette situation que nous prenons en compte le pronom personnel sujet *jo, je*, qui a bien une voyelle flottante et est capable de s'élider, mais qui commence son parcours en t.a.fr. en ne s'élidant jamais. Il s'agit alors de considérer que le pronom sujet, dont l'expression n'est pas encore obligatoire (MOIGNET 2002 [1976], p. 127-129; MARCHELLO-NIZIA et PRÉVOST 2020, GGHF, p. 1055-1070), a la structure détachée d'un pronom sujet emphatique : il ne s'inscrit pas dans une structure PROCS + VER solidaire, mais résulte de l'insertion, autonome, d'un pronom devant le verbe (14.50).

(14.50) DOM = [MOTFONC]

Lorsque le mot fonctionnel est le seul composant de son nœud syntaxique et que celui-ci est suffisamment indépendant du nœud syntaxique subséquent, le mot fonctionnel peut former son propre domaine.

p.ex. *jo atendi quet* (*Alexis*, v. 389, 479)²⁶

Dans l'exemple donné en (14.50), nous supposons donc que *jo* forme son propre domaine computationnel, distinct de celui de *atendi*. C'est le caractère facultatif de l'expression du sujet qui rend possible cette structure ; *atendi quet* est possible sans *jo* lorsque la langue est encore *pro drop*. La coïncidence entre la montée en puissance de l'expression du pronom CS et des formes élidées de *je* est traitée *infra* (sec. 14.3.2.2). Notre analyse en termes de proximité syntaxique rejoint les termes de MOIGNET (1965), qui parle de séries *disjointes* et *conjointes* pour les pronoms, plutôt que de formes fortes et faibles. Le caractère fort ou faible du pronom sujet n'est donc que la conséquence de sa disjonction ou jonction syntaxique et, par dérivation, phonologique, avec le verbe.

14.3.1.2 Condition de taille

Les conditions syntaxiques énoncées *supra* ne suffisent pas à définir le domaine de l'élision de la période 1. En effet, nous avons avancé en sec. 14.2.2 que les voyelles finales des mots polysyllabiques fonctionnels ou lexicaux (dans ce cas paroxytons) étaient également atteintes par le flottement, sous l'effet d'un processus distinct (COND.ə). Or nos données graphiques, que nous considérons comme les témoins de l'élision de la première période, témoignent de deux faits qui compliquent la situation. D'une part, les mots fonctionnels polysyllabiques ne sont pas (sauf collocation) atteints par l'élision graphique de la période 1 : *une, ele* et, moyennant l'affaiblissement de sa distribution syntaxique complémentaire, *come*, ne s'élident pas, ou marginalement. D'autre part, l'élision dans les collocations dès *Roland* (*ensembl od*),²⁷ alors même que la réduction des hiatus internes n'est pas encore attestée, indique que la voyelle finale de l'ADV *ensemble* est déjà flottante, tandis que le processus est retardé en position interne.

Ce second élément fait que l'on ne peut se contenter de supposer que COND.ə n'est pas encore active et que seuls les monosyllabes fonctionnels de structure archaïque CV disposent de voyelles finales flottantes. Au contraire, dès *Roland*, il est probable que tout schwa final est déjà flottant. Dans ces conditions, il convient de considérer que la raison pour laquelle *une, ele* et *come*, ainsi que *entre* et *encontre* hors collocation, ne s'élident

25: « 'degrees of connectedness' of words in sequence that are appropriate for governing the applicability of external sandhi rules. » (SELKIRK 1984, p. 302)

26: Cité par MARCHELLO-NIZIA (2015) ; la métrique d'*Alexis* ne suppose jamais d'élision de *je*.

pas en période 1 est liée à leur taille : tous les mots qui s'élident hors collocation en période 1 sont des monosyllabes. Nous proposons donc de réécrire $\text{DOM} = [\text{MOT}_{\text{FONC}} \text{MOT}_{\text{LEX}}]$ (14.49) *supra* en intégrant une contrainte sur la taille du mot fonctionnel éligible au statut de clitique, tel qu'en (14.51).

- (14.51) $\text{DOM} = [\text{MOT}_{\text{FONC}}(\sigma) \text{MOT}_{\text{LEX}}]$
 Un mot fonctionnel peut appartenir au domaine du mot lexical suivant lorsqu'il a un nœud mère avec celui-ci de niveau suffisamment bas dans l'arbre syntaxique, et qu'il ne contient qu'une seule syllabe.

Le corrolaire de $\text{DOM} = [\text{MOT}_{\text{FONC}}(\sigma) \text{MOT}_{\text{LEX}}]$ (14.51) est que tout mot polysyllabique, en période 1, forme son propre domaine computationnel. On peut donc également écrire une sous-version de la description $\text{DOM} = [\text{MOT}_{\text{FONC}}]$ (14.50) *supra*. Ainsi, selon l'énoncé en (14.52), tout mot polysyllabique correspond nécessairement à un domaine computationnel en période 1, que le mot soit fonctionnel ou lexical.

- (14.52) $[\text{MOT}(\sigma\sigma\dots)] \Rightarrow \text{DOM}$
 Lorsqu'un mot est polysyllabique, il doit correspondre à son propre domaine computationnel.

Ainsi, en période 1, ne sont clitiques que les mots fonctionnels monosyllabiques qui entretiennent un rapport syntaxique étroit avec le mot lexical le plus proche sur leur droite. Les mots lexicaux et les mots fonctionnels polysyllabiques, eux, doivent constituer leur propre domaine.

14.3.1.3 En somme

La situation de la période 1 de l'a.fr., où ne s'élident que certains monosyllabes fonctionnels dont la voyelle finale est flottante, correspond à ce qu'il est convenu de qualifier de cliticité. La cliticité concerne un mot fonctionnel qui, à partir de la structure syntaxique, est placé dans le même domaine computationnel que le mot lexical le plus proche, sur sa droite pour ce qui est de l'élision. À cette condition syntaxique s'ajoute, dans le cas de la première période de l'a.fr., une condition de taille : seuls les monosyllabes fonctionnels peuvent être clitiques. Tous les autres mots doivent être insérés dans leur propre domaine. Les critères qui expliquent la répartition des formes sont reportés sur le diagramme en fig. 14.3, reproduction annotée de celui présenté en début de chapitre.

27: Cf. sec. 7.2.3, 7.2.3.4, 7.3.1, 12.3.2.1.

14.3.2 Période 2 : élision au sein d'un domaine plus large

Les grammairiens du xvi^e s., de concert avec les données métriques, témoignent d'une élision dont le domaine est bien plus large que celui qui unit les clitiques de structure archaïque CV et le mot lexical le plus proche sur leur droite. Telle que définie en sec. 13.2.3, l'élision métrique concerne l'hémistiche ou le vers — et l'hémistiche, en tant que domaine spécifié d'élision (coupe épique), cède progressivement la place au vers au xiv^e s. Dans le témoignage des grammairiens du xvi^e s., le domaine de l'élision n'est guère désigné que par son terme : l'élision ne se pratique

pas devant ponctuation ou pause (sec. 13.3.1). Dans les deux cas, l'élision se pratique donc au sein d'une unité qui peut comprendre non seulement des mots fonctionnels de structure archaïque non CV, mais également plusieurs mots lexicaux.

Nous avons avancé en sec. 13.4.1 que la transition entre l'état dont témoignent nos données graphiques (période 1), et celui dont témoignent les grammairiens du xvi^e s. et la métrique (période 2) est l'élargissement général d'un domaine computationnel de l'a.fr., qui aurait eu lieu dans la seconde moitié du xii^e s. et la première moitié du xiii^e s. Nous avons relevé que cet élargissement du domaine computationnel était souvent analysé en termes d'élargissement du groupe accentuel, d'affaiblissement des phénomènes d'enclise et des consonnes finales, et d'instauration du système de la liaison consonantique. Tous ces phénomènes relèvent de propriétés qui regroupent les mots en unités plus grandes. Aussi, nous avons proposé de considérer que l'élargissement du domaine de l'élision s'inscrivait dans ce même mouvement général, et que le domaine de l'élision était, selon toute vraisemblance, également celui de l'accent et du sandhi consonantique.

L'identification précise du domaine de l'élision n'est pas chose aisée. La métrique, par définition, ne propose qu'un témoignage indirect, s'il n'est suspect. Nous avons proposé de considérer, en sec. 13.3.2, que l'addition de la grammaire d'association et de la paraphonologie métrique forçaient le domaine computationnel à correspondre à un domaine métrique. Aussi l'analyse du domaine de l'élision en métrique ne saurait être une preuve suffisante. Quant au témoignage des grammairiens, il est postérieur de plusieurs siècles au changement supposé, et ne témoigne que d'un acrolect normatif, voire de plusieurs variétés acrolectales (cf. débat entre Péletier et Meigret, p. 13.3.1). Par ailleurs, les faits morphosyntaxiques sont eux-mêmes voilés par la grammaticalisation : certaines enclises grammaticalisées sont parvenues jusqu'au fr.mod. (*del* → *du* ; *al* → *au*). C'est seulement la fin de productivité qui peut être fiable, et celle-ci est pour partie bien antérieure à l'époque visée : certaines enclises disparaissent dès le début de l'a.fr.²⁸ Enfin, les éléments de sandhi consonantique ne sont pas plus explicites, dans la mesure où ils ne sont qu'en partie, ou avec délai, transcrits dans la graphie ; comme l'élision, l'ampleur et la chronologie précise du phénomène nous échappent nécessairement.

Certains éléments, dans nos données, permettent néanmoins de préciser quelque peu la façon dont le domaine computationnel de la période 1 s'élargit pour aboutir à celui de la période 2. Comme avancé en sec. 14.3.1, nous envisageons la définition du domaine computationnel comme le cumul de conditions syntaxiques et non syntaxiques ; le fait que des contraintes non syntaxiques soient en jeu explique la non-isomorphie observée entre certains domaines computationnels et les constituants syntaxiques disponibles. Ce non-isomorphisme est l'un des arguments clef qui soutient la phonologie prosodique : la syntaxe n'est pas suffisante pour prédire l'empan de certains processus phonologiques. La phonologie prosodique classique, pourtant, ne parvient que difficilement à la prédiction ; certains de ses domaines peuvent varier en fonction du locuteur, du registre, du débit, etc. C'est le cas des deux constituants prosodiques qui seraient de bons candidats pour représenter le domaine de l'élision : le syntagme phonologique et le syntagme intonatif.

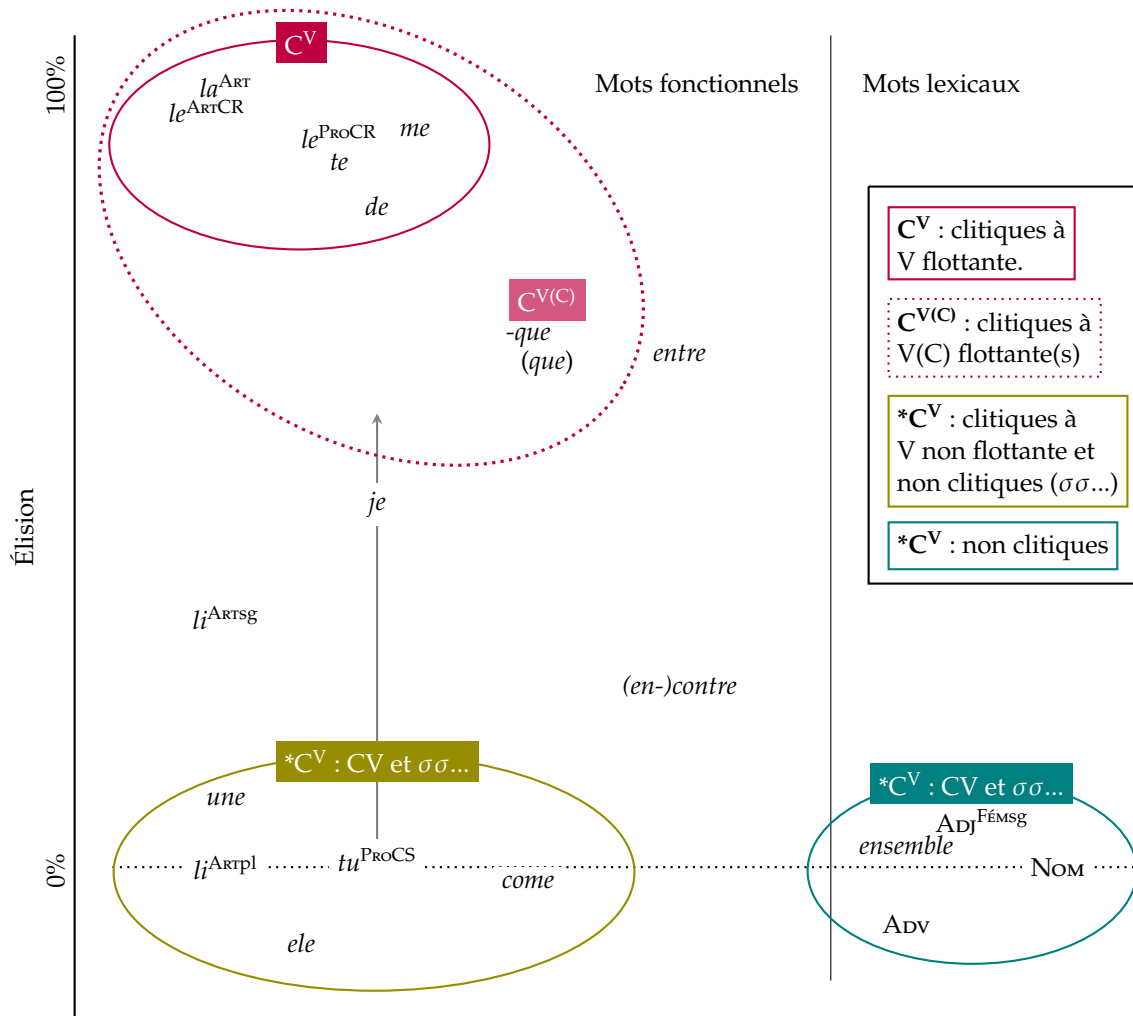


FIGURE 14.3 – Distribution des taux d’élision en fonction des critères relevés. Un taux d’élision négatif signifie qu’on trouve plus de défauts devant #C que devant #V. Les valeurs des ordonnées sont approximatives. Un élément en exposant correspond à une mélodie flottante; « σσ... » à un mot polysyllabique.

14.3.2.1 Un potentiel candidat : le syntagme phonologique φ

Les travaux de RAINSFORD (2010, 2011a,b) conduisent à considérer que le syntagme phonologique (φ) serait un bon candidat pour représenter le domaine de l’élision, de la liaison et de l’accentuation. Le syntagme phonologique est défini comme contenant une tête lexicale et tous les mots contenus dans sa projection maximale sur son côté non récursif (ici, à gauche; Post 2000, p. 34), c’est-à-dire comme un groupe clitique CG₁ et tous les groupes clitiques sur son côté non récursif jusqu’au premier groupe clitique situé hors de la projection maximale de CG₁ (NESPOR et VOGEL 2007 [1986], p. 168). Sous cette définition, une variété de φ peut être générée, de la tête lexicale isolée à des configurations comme DÉT + ADJ + NOM (14.53 a), en passant par DÉT + NOM (14.53 b) et PRÉP + PROCR + VER (14.53 c) ou PROCS + VER (14.53 d).

28: P.ex. *poros* pour *poro se*, *semprel* pour *sempre lo* (Eulalie), *sis* pour *si+se* (Jonas), *porqueim* pour *porquei me* (Alexis), *sim* pour *si+me*, *kis* pour *ki se* (Roland); NYROP (1899, p. 239-240).

(14.53) Exemple de syntagmes phonologiques

- | | |
|--------------------------------|----------------|
| (a) <i>une samblable chose</i> | (a)lea, p. 13) |
| (b) <i>les temptacions</i> | (abe, p. 1) |
| (c) <i>pour li faire</i> | (abe, p. 23) |
| (d) <i>j avoye</i> | (abe, p. 25) |

Post (2000). *Tonal and Phrasal Structures in French Intonation*.

ailea : RAOUL DE HOUDENC (1207*). *Le roman des ailes*. Ms. A ; Dees : Yonne ; compo. pic., ca 1207 ; ms. frc., ca 1288 ; MS (RAOUL DE HOUDENC 1983).

En reprenant les définitions de domaine volontairement naïves données en sec. 14.3.1, un syntagme phonologique comprend un nœud lexical et tous les nœuds, lexicaux ou fonctionnels, qui ont un nœud commun avec lui, et dont il est la tête. Ces syntagmes phonologiques peuvent donc, *a priori*, être dérivés directement depuis la syntaxe : selon ces définitions, il y a un constituant syntaxique isomorphe avec le domaine phonologique. Mais le syntagme phonologique, tout comme les constituants supérieurs de la hiérarchie prosodique, peut aussi être restructuré : si ADJ + NOM est obligatoirement un syntagme phonologique, NOM + ADJ ne l'est qu'optionnellement. En effet, sa structure fondamentale est [... NOM]_φ [ADJ]_φ, mais ces deux φ peuvent fusionner pour former [NOM ADJ]_φ.

Mécanismes de restructuration de φ. L'exemple devenu classique de NESPOR et VOGEL (2007 [1986], p. 172-173) pour le procédé de restructuration des syntagmes phonologiques est le *raddoppiamento sintattico* (RS) de l'italien florentin. Ce processus affecte la consonne initiale d'un mot ω₂ en la géminant lorsque (a) celle-ci est suivie d'une sonante non nasale, que (b) le mot précédent ω₁ est oxyton, et (c) que ω₁ et ω₂ sont placés dans le même syntagme phonologique (14.54; *id.* p. 165-166).

abe : JEAN DE MEUN (1288*). *Traduction de la première épître de P. Abélard*. Dees : région parisienne ; compo. frc., ca 1280 ; ms. Paris, ca 1395 ; MS3 (JEAN DE MEUN 1934).

(14.54) C → C : / [[... 'σ]_{ω1} [___ x^[son. - nas.] ...]_{ω2} ...]_φ

NESPOR et al. (2007 [1986]). *Prosodic Phonology*.

Ainsi, le RS s'applique dans une structure qui forme nécessairement un syntagme phonologique, comme DÉT + ADJ (+ NOM) en (14.55). Dans une structure sujette à restructuration comme NOM + ADJ, le RS ne s'applique que si la restructuration a lieu : il ne s'applique pas si le NOM et l'ADJ demeurent chacun dans leur syntagme phonologique (14.56), mais s'applique si les deux constituants ont été fusionnés (14.57).

- | | | |
|-------------------------|-----------------------------------------------------------------------|-----------|
| (14.55) DÉT + ADJ + NOM | [DÉT ADJ NOM] _φ
<i>tré p:iccoli cobra</i> | RS : [p:] |
| (14.56) DÉT + NOM + ADJ | [DÉT NOM] _φ [ADJ] _φ
<i>I caribú nani</i> | *RS : [n] |
| (14.57) DÉT + NOM + ADJ | [DÉT NOM ADJ] _φ
<i>I caribú n:ani</i> | RS : [n:] |

Adapté de NESPOR et VOGEL (2007 [1986], p. 170).

Dans la phrase *Era venuto con tre piccoli cobra*, « Il est venu avec trois petit cobras ».

Problème heuristique de la restructuration de φ. Si NESPOR et VOGEL (2007 [1986], p. 179) postulent que la restructuration des syntagmes phonologiques n'est pas active en français, la restructuration semble nécessaire pour rendre compte de multiples phénomènes phonologiques en français, ainsi que RAINSFORD (2011b, p. 21) le déduit des travaux de POST (2000) et d'autres. Ce processus de restructuration pose problème dans la démarche historique : la restructuration est un outil descriptif efficace, mais limite fortement la puissance prédictive du modèle. Par ailleurs, de nombreux facteurs tiers sont à l'œuvre, comme le débit (NESPOR et VOGEL 2007 [1986], p. 173-174; POST 2000) ou la taille (en nombre de syllabes) des φ à fusionner (POST 2000). De ce fait, la structure

prosodique d'un énoncé donné ne peut pas être établie *a priori* : elle est induite de l'empan des phénomènes phonologiques constatés. On ne peut donc déduire, en diachronie, l'existence d'un phénomène à partir de la structure prosodique, au risque d'une circularité néfaste — c'est la raison pour laquelle RAINSFORD (2011a, p. 270-272) s'en tient à traiter d'*unités accentuelles possibles*. De fait, même si l'on acceptait de recourir aux syntagmes phonologiques, ceux-ci seraient définis comme étant variables et cette variation serait impossible à mesurer dans nos données.

14.3.2.2 Un candidat fiable, mais insuffisant : la syntaxe

La seule chose qui soit fiable, c'est donc la structure syntaxique, et il faut néanmoins admettre que la transcription de la structure syntaxique en domaines phonologiques ne résulte pas nécessairement en une isomorphie entre les deux structures. Si c'est ce non-isomorphisme entre structure syntaxique et domaines phonologiques qui est l'argument principal en faveur de l'existence de la phonologie prosodique (NESPOR et VOGEL 2007 [1986], p. 27-60, *et passim*),²⁹ cet argument se révèle donc problématique dans notre cas précis.

De ce fait, plutôt que d'essayer de définir virtuellement quels sont les constituants syntaxiques qui sont transcrits, en phonologie, en tant que domaines d'élision, nous nous en tenons à les traiter dans l'abstraction. Nous ne cherchons donc pas à déterminer exhaustivement les contextes par lesquels le domaine computationnel s'élargit. On peut, néanmoins, formuler quelques remarques, qui réduisent le niveau d'abstraction par lequel nous accédons à la relation entre syntaxe et domaines computationnels.

Quelques adjectifs élidés. Les données au chap. 5 concernant les adjectifs féminins ont permis d'isoler certains cas où la structure ADJ + NOM semblait particulièrement propice à l'élision. Nous avons relevé en sec. 5.2.3.2 les cas de *seint* + NOM (*seint eglise, seint agnes, seint escripture*), *long* + NOM (*long aleine, long espee*) et *mal* + NOM (*mal heure, mal estrine, mal aventure, mal entente, mal envie, mal uxore*). Si certains de ces cas semblent tenir de la locution plus ou moins grammaticalisée (*mal heure, long aleine*) et non de l'élision en tant que telle, ce n'est probablement pas le cas de l'ensemble de ces occurrences : la liberté paradigmatique semble complète pour *seint*, probablement aussi pour *long* (*long espee* n'est pas enregistré en tant que locution), tandis que la forte productivité de *mal* pour les locutions et la préfixation rendent difficile l'interprétation de ces résultats. Nous avons également noté, à la même section, que c'était dans le même contexte ADJ + NOM que SÉGÉRAL et SCHEER (2020, GGHF, p. 407) relevaient un cas précoce de sandhi externe, avec l'occurrence *sein Gabriel* au lieu de *seint Gabriel* (*Roland*, v. 2847).

Ces élisions, rares et minoritaires, dans le contexte ADJ + NOM peuvent donc être considérées comme les premières manifestations d'un domaine computationnel qui, parfois (dimension stochastique de l'élargissement du domaine), inclut l'adjectif antéposé dans le domaine computationnel du nom. Dans ce cas, ces premières attestations remontent aussi loin que *Roland*. Surtout, elles culminent à la fin du XII^e s. et au début du

Dans la phrase *I caribú nani sono estinti*,
« les caribus nains sont éteints ».

29: C'est l'une des différences entre les approches métriques comme celles de Liberman et Prince et les approches de phonologie prosodique classique : « If our claims about the existence and nature of ω are well-founded, one cannot conceive of prosodic structure as being a mere replication of syntactic structure, as is done by Liberman and Prince. » (SELKIRK 1981, p. 130).

xiii^e s., c'est-à-dire au cœur de la période identifiée pour l'élargissement du domaine computationnel (cf. fig. 5.10 en sec. 5.2.4, dont est extraite la figure marginale 14.4). Certes, la majorité des ces occurrences se trouve en anglo-normand (cf. sec. 5.3.5), mais cela peut vouloir dire que les habitudes graphiques anglo-normandes sont moins sensibles à la norme (établissement d'un code graphique qui ne note plus les nouvelles élisions) que ne le sont les habitudes graphiques continentales.³⁰

À l'opposé des structures ADJ + NOM, les structures NOM + ADJ ne sont pas attestées dans nos données, étant donné que les élisions graphiques de noms sont nulles dans notre corpus. Il semble donc bien que l'élargissement du domaine procède d'abord par les structures ADJ + NOM avant d'atteindre éventuellement les structures NOM + ADJ; si celles-ci sont jamais atteintes, hors métrique. Ce fait, ici, est cohérent avec la restructuration des syntagmes phonologiques, et donc avec les données observées dans de multiples langues dans le champ de la phonologie prosodique. Ces faits peuvent être réduits aux énoncés en (14.58) et (14.59).

(14.58) On peut interpréter les quelques occurrences d'élision dans le contexte ADJ + NOM comme les traces marginales de l'élargissement du domaine computationnel à ce domaine.

(14.59) Il y a des traces de l'élision pour ADJ + NOM mais pas pour NOM + ADJ.

Le chemin de *jo*, une fenêtre sur l'élargissement du domaine. Tel qu'il est analysé par MARCHELLO-NIZIA (2015), le pronom personnel sujet *jo*³¹ commence par exister dans la seule série des formes fortes, avant que des formes faibles, sujettes à l'élision, n'apparaissent. Couplé avec l'évolution d'un système où l'expression du pronom personnel sujet est facultative et emphatique vers un système où l'expression de ce pronom est obligatoire et non marquée, l'apparition des premières élisions du pronom semble donc témoigner de l'élargissement de ce domaine computationnel. Cet élargissement est peut-être déjà à l'œuvre dans *Roland* (élisions métriques et non graphiques) et commence à être ponctuellement transcrit peu après; l'élision graphique va croissante jusqu'au début du xiii^e s., avant de se stabiliser. MARCHELLO-NIZIA (2015) note dans le détail que cet élargissement se fait contexte syntaxique par contexte syntaxique; la dimension stochastique de l'élargissement est donc réduite.

Ainsi, comme relevé en sec. 14.3.1.1 *supra*, *jo* n'est jamais élidé dans *Alexis*. À cette époque, l'expression du pronom personnel sujet est facultative (MOIGNET 2002 [1976], p. 127-129; MARCHELLO-NIZIA et PRÉVOST 2020, p. 1055-1070) et peut-être emphatique; dans ce cas, le pronom personnel sujet est détaché du verbe. De ce fait, il peut se retrouver sans domaine auquel appartenir, or tout matériel phonologique doit appartenir à un domaine computationnel pour être traité par la phonologie. Dans ce cas, le mot forme seul son domaine, et l'élision est proscrite, comme illustré en (14.50) *supra*, reproduit ici :

(14.50) DOM = [MOTFONC]

Lorsque le mot fonctionnel est le seul composant de son nœud syntaxique et que celui-ci est suffisamment indépendant du

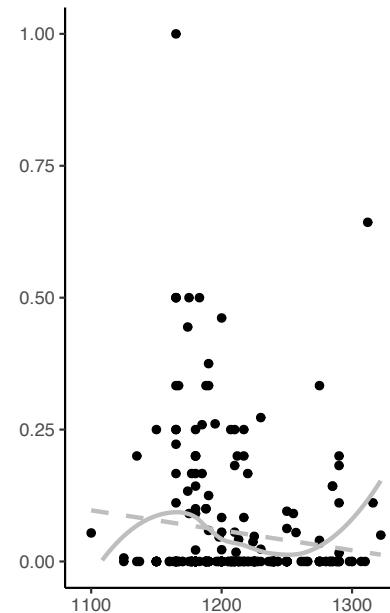


FIGURE 14.4 – Taux d'élision des adjectifs féminins singuliers. Extrait de la fig. 5.10.b.

30: Cf. sec. 13.4.2.

nœud syntaxique subséquent, le mot fonctionnel peut former son propre domaine.

p.ex. *jo atendi quet* (*Alexis*, v. 389, 479)

La montée en puissance de l'élision de *jo* traduit donc deux phénomènes : d'une part le fait que l'expression du pronom sujet est de plus en plus obligatoire, d'autre part le fait que lorsque le pronom est exprimé dans son rôle par défaut, il appartient au domaine computationnel du verbe lorsque le verbe le suit immédiatement. MARCHELLO-NIZIA (2015) note qu'à ces contextes PRO + VER (attestés dès *Roland* mais uniquement par la métrique) s'ajoutent ensuite les contextes PRO + *en* + VER au milieu du XII^e s. (*Énéas*), puis PRO + *y* + VER, des contextes de proposition intercalée (*se j'amender le dei*, « si je dois l'amender », *Becket*, v. 4410) et de postposition du pronom VER + PRO (*ne voil j'en ço*, « je ne veux, en cela », *Becket*, v. 3434). Le constat de l'autrice que la situation se stabilise au XIII^e s. est corroboré dans nos données et est directement visible par la régression locale en figure marginale 14.5. Pour le pronom personnel sujet, on peut donc suivre par le menu l'élargissement du domaine computationnel en termes syntaxiques, tel qu'énoncé en (14.60 - 14.61).

(14.60) Le pronom personnel sujet est d'abord dans un domaine computationnel autonome. On peut suivre son inclusion progressive dans un domaine de plus en plus large, qui va de pair avec son expression de moins en moins facultative

(14.61) Élargissement du domaine de l'élision du PRO_{PERSCS} :

- | | |
|-----------------------------------------------|--------------------------|
| .i. PRO + VER | déb. XII ^e s. |
| .ii. PRO + <i>en</i> + VER | m. XII ^e s. |
| .iii. PRO + <i>y</i> + VER | fin XII ^e s. |
| PRO + [... [̣]] + VER ³² | |
| VER + PRO | |

L'énoncé en (14.60), qui porte sur tous les pronoms sujets, ne pose pas de difficulté de sur-génération, dans la mesure où le pronom personnel sujet *je*, *jo* (etc.) est le seul pronom personnel au CS à s'élider. Le pronom personnel de deuxième personne *tu*, *te* et le pronom personnel de troisième personne *ele* y échappent, puisqu'ils ne répondent pas à la règle COND_{CV}, qui seule autorise l'élision graphique (sec. 14.2.1.2). Les pronoms personnels régimes, en revanche, du moment qu'ils répondent à COND_{CV}, s'élident systématiquement ; cela est naturel étant donné que la position syntaxique du pronom CR est distincte de celle du pronom CS.

Le pronom sujet dispose donc d'un statut ambigu au sein des catégories que nous avons tracées, ambiguïté représentée par la flèche sur les graphes en fig. 14.1 et 14.3 *supra*. Pour mieux comprendre sa situation, il convient de ne pas considérer le domaine du pronom, mais celui du verbe, puisque c'est la perte du domaine autonome qui déclenche l'élision. En reprenant les principes d'établissement du domaine computationnel utilisées en sec. 14.3.1, on peut donc avancer que le domaine computationnel du pronom personnel disparaît à mesure qu'il est intégré à celui du verbe, d'abord dans les contextes où rien ne les sépare (14.62 b), puis en admettant progressivement de plus en plus d'éléments interjetés entre eux (14.62 c) et finalement la possibilité que le pronom se trouve à gauche du verbe (14.62 d).

31: Ici, nous désignons par *jo* l'ensemble des formes qui précèdent la séparation entre les séries faibles et fortes ainsi que l'ensemble des formes faibles du pronom : *jo*, *jeo*, *jou*, *je* et *g*. Cf. table 9.1.

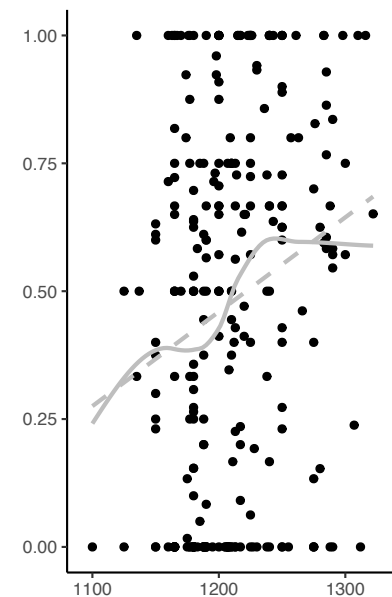


FIGURE 14.5 – Taux d'élision du pronom personnel sujet *jo*. Extrait de la fig. 9.19.b.

(14.62) Élargissement progressif du domaine du verbe

- (a) $\text{DOM} \Rightarrow [\text{VER}]$ t.a.fr.
 donc $\text{PRO}^{33} \Leftarrow \text{DOM}$ (14.50)
- (b) $\text{DOM} \Rightarrow [\text{PRO VER}]$ déb. XII^e s.
- (c) $\text{DOM} \Rightarrow [\text{PRO MOTFONC VER}]$ m.-fin XII^e s.
- (d) $\text{DOM} \Rightarrow [\text{PRO } [\dots^{\text{r}}] \text{VER}]$
- (e) $\text{DOM} \Rightarrow [(\text{PRO } [\dots^{\text{r}}]) \text{VER} (\text{PRO})]$

32: [...^r]: syntagme intercalé par mouvement syntaxique.

En effet, *jo* témoigne de l'élargissement progressif du domaine, mais il s'agit d'un monosyllabe fonctionnel. Nous avons postulé en sec. 14.3.1.1 que *jo* commençait son existence, en t.a.fr, isolé dans son propre domaine (configuration 14.50), et que son élision résultait uniquement de l'élargissement du domaine : dès que *jo* est inséré dans le domaine du mot qui le suit, il est sujet à l'élision.

(14.63) *jo* lorsque $\text{DOM} \Rightarrow [\text{VER}]$ (14.62 a)

C V
 | ↑
 ʒ o

(14.64) *j ai* lorsque $\text{DOM} \Rightarrow [\text{PRO VER}]$ (14.62 b)

C V C V
 | |
 ʒ o ε

(14.65) *j en ai*³⁴ lorsque $\text{DOM} \Rightarrow [\text{PRO en VER}]$ (14.62 c)

C V C V C V
 | | |
 ʒ o e n ε ...s livres

(14.66) *j es livres*³⁵ lorsque $\text{DOM} \Rightarrow [\text{PRO } [\dots^{\text{r}}] \text{VER}]$ (14.62 d)

C V C V C V
 | | |
 ʒ o e s

33: Ici, PRO abrège PRO_{PERS}CS.

34: Par exemple maur p. 20 : *fait la danzele 'beaus amis / sacheiz j en ai bien conseil pris*, « la demoiselle dit : sachez, mon bel ami, que j'ai bien retenu [ce] conseil » (trad. pers.).

maur : BENOÎT DE SAINTE-MAURE (1170*). *Le roman de Troie*. Dees : [Vendee, Deux-Sevres]; compo. poit., ca 1170; ms. a.n., ca 1210; CR3 (BENOÎT DE SAINTE-MAURE 1904); DEAF (ToDo) : BenTroieC.

En somme. Si les possibilités de non-isomorphie entre syntaxe et domaine computationnel rendent impossible toute prédiction théorique lorsque les données sont silencieuses, nous sommes en mesure de documenter certains détails de l'élargissement du domaine computationnel. Ainsi, nos données semblent témoigner, au moment précis où est supposé se dérouler le changement d'unité accentuelle, d'une contamination des contextes ADJ + NOM mais pas des contextes NOM + ADJ. De même, l'élargissement progressif des contextes d'élision du pronom personnel sujet *jo* laisse supposer que le domaine acceptant l'élision grandit progressivement, en incorporant de plus en plus d'éléments interjetés entre le pronom et le verbe.

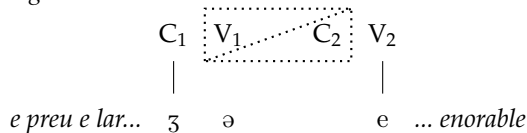
14.3.3 Fonctionnement phonologique synchronique

Au sein d'un domaine computationnel donné, l'élision métrique et celle, hors métrique, de la période 2, reposent probablement sur les mêmes

mécanismes que celle documentée en période 1 pour les clitiques. En effet, selon les témoignages métriques et ceux des grammairiens, l'élision n'a pas lieu en fin de domaine, mais a lieu partout à l'intérieur de celui-ci. Il ne s'agit donc pas de chercher à mettre en place un processus qui viserait spécifiquement les schwas finaux, dans la mesure où les schwas finaux de domaines et les schwas préconsonantiques sont protégés de l'élision. C'est pourquoi nous avons proposé de considérer, dès la période 1, que tout schwa final était flottant; si, en période 1, l'élision de ces schwas finaux de polysyllabes ne se voit pas, c'est que ces mots sont encore placés dans des domaines computationnels séparés. L'élargissement du domaine de l'élision des polysyllabes est donc en tout point semblable à celui du pronom sujet *jo* : ils sont prêts à s'élider (leur voyelle finale est flottante) et n'attendent qu'une chose pour le faire : se trouver dans un domaine dont ils ne sont pas le dernier élément.

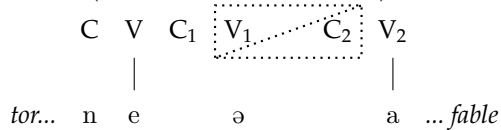
Si l'on admet que le flottement est d'ores-et-déjà en place, ce qu'indiquent également les cas précoces de collocations comme *ensembl od* dans *Roland*, l'élision découle naturellement, comme en période 1, de *VCVIDES. Nous utilisons l'élision métrique pour illustrer ce fonctionnement, d'abord en (14.67), sans représenter le lien entre le domaine computationnel imposé par la métrique et les représentations.

(14.67) *large e*³⁶

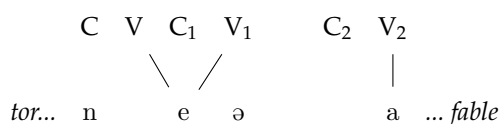


Lorsqu'un mot à hiatus Və final est placé devant initiale vocalique, il est impossible de prédire qui, de l'élision ou de la réduction du hiatus, l'emporte; notre mécanisme combinant la mélodie flottante (nécessaire pour la réduction du hiatus comme pour l'élision) et la non disponibilité du VC vide prédit dans tous les cas que si le schwa final demeure en hiatus avec la voyelle précédente, il ne peut demeurer en hiatus avec la suivante (14.68), et ce même si le hiatus Və est résolu par l'allongement de V (14.69).

(14.68) *tornee a* (sans réduction du hiatus)³⁷



(14.69) *tornee a* (avec réduction du hiatus et allongement préservé)



Peu importe le domaine qu'on lui fournit, ce mécanisme fonctionne : une voyelle finale flottante placée devant initiale vocalique est supprimée par la disparition de la position vocalique synchrone qui aurait pu l'accueillir devant initiale consonantique ou en fin de domaine.

En fin de domaine, comme en période 1, les voyelles finales flottantes se maintiennent : le domaine dans lequel est bornée la phonologie ne

35: Par exemple benoit p. 7 : *si com j es libres ai trové*, « ainsi que j'ai trouvé en ces livres » (trad. pers.). Nous représentons l'enclise *en+les* → *es* comme lexicalisée, puisque ce n'est pas notre propos.

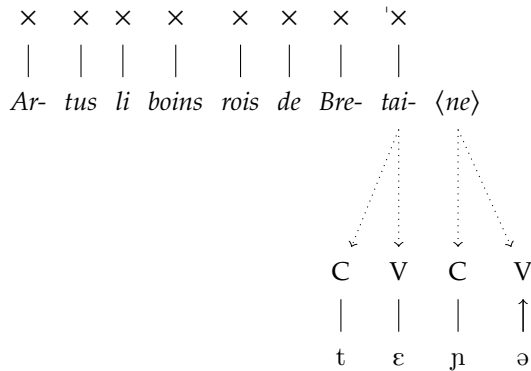
benoit : BENOIT DE SAINTE-MAURE (1174*). *Chronique des ducs de Normandie*. Ms. T; Dees : Vendée, Deux-Sèvres; compo. poit., ca 1174; ms. poit., ca 1190; MS1 (BENOIT DE SAINTE-MAURE 1951-1967)

36: Par exemple *Yvain* (yvh) v. 23 : *et preu et large et enorable*, « preux, généreux et honnêtes » (trad. WALTER in CHRÉTIEN DE TROYES 1994).

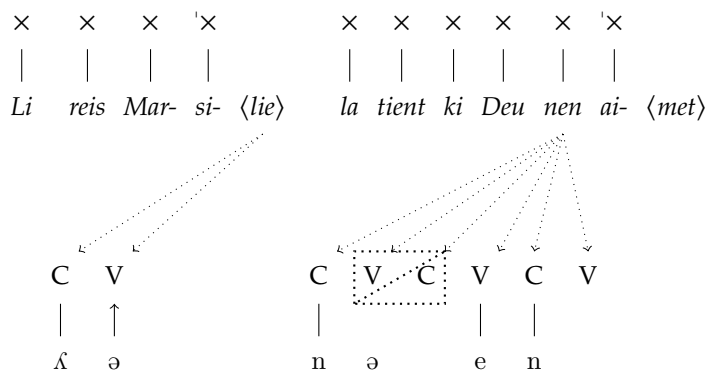
yvh : CHRÉTIEN DE TROYES (1177*). *Le Chevalier au lion*. Ms. H; Dees : nil; compo. champ. mérid., ca 1177; ms. champ., ca 1235; MS (transcr. inédite R. L. H. Lops); également contenu dans chret2.

dispose d'aucun VC vide et AssocV peut donc s'appliquer. C'est le cas, par exemple, en terminaison de vers (14.70) ou en coupe épique (14.71). Sur ces schémas, les flèches en pointillés donnent la correspondance entre les syllabes, telles qu'alignées sur le gabarit métrique, et leur représentation CVCV ; elles ne suggèrent en aucun cas un plan représentationnel supplémentaire.

(14.70) *Bretaigne* en terminaison féminine³⁸



(14.71) *Marsilie* en coupe épique, *ne+en* élidé, *aimet* en term. fém.³⁹



Ainsi, sur *Bretaigne* en terminaison de vers (14.70) et sur *Marsilie* (paroxyton trisyllabique)⁴⁰ en coupe épique (14.71), puisque la dernière position métrique à laquelle les mots sont associés est proéminente, la computation est insensible au contexte subséquent. La situation est donc toujours la même, peu importe que ces paroxytons soient dotés d'une consonne finale ou non (*Bretaigne* est traité comme *aimet*), et qu'ils soient 'suivis' d'une initiale vocalique ou consonnantique : du point de vue computationnel, ils ne sont suivis de rien.

En (14.71), on peut également observer une élision probablement normale en période 1 : *ne en* est réduit à *nen*. Au sein du domaine métrique, les élisions sont la règle, qu'elles concernent des clitiques ou non (sec. 13.2.3) ; un exemple est donné en (14.72) pour un mot fonctionnel non clitique (polysyllabique).

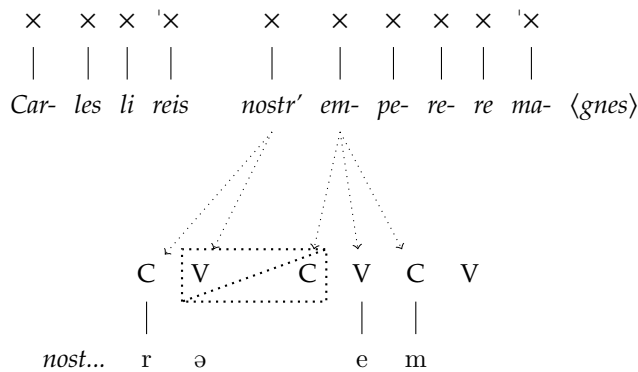
(14.72) *Nostre* élidé au sein d'un hémistiche⁴¹

37: Par exemple *Yvain* (yvh) v. 24 : *or est amors torne a fable*, « la réputation d'Amour en est fort amoindrie » (trad. *id.*)

38: *Yvain* (yvh), v. 1 : « Arthur, le bon roi de Bretagne » (trad. WALTER in CHRÉTIEN DE TROYES 1994).

39: *Roland*, v. 7 : « Le roi Marsile la tient, qui n'aime pas dieu » (trad. SHORT 1990).

40: Les graphies et <ni> non syllabiques et posttoniques sont courantes dans *Roland* (SHORT 2007, p. 101, 103), alors que, dès Philippe de Thaon, l'a.n. semble avoir perdu les oppositions /l/ ~ /ʎ/ et /n/ ~ /ɲ/ (SHORT 2007, p. 100, 102). Les graphies <ri> posttoniques pour <r> compliquent encore un peu plus la situation ; elles sont parfois métathétiques (*glorie, estorie*) (SHORT 2007, p. 104). Dans tous les cas, ce <i> posttonique n'est jamais syllabique. La neutralisation du trait palatal étant attestée seulement après *Roland*, nous transcrivons *-lie* de *Marsilie* /ʎə/.



14.3.4 En somme

Il est difficile d'établir avec précision les modalités de l'élargissement du domaine computationnel, y compris sa taille finale. Nos occurrences d'adjectifs féminins singuliers en défaut de <e> dans des structures ADJ + NOM coïncident avec le centre de la période du changement de domaine, ce qui plaide pour que ce contexte fût parmi les premiers à être atteint par l'élargissement. Les données de MARCHELLO-NIZIA (2015) sur *jo*, corroborées par nos observations, permettent d'établir un peu plus précisément l'élargissement du domaine computationnel du vers, ce qui place *jo* dans une posture paradoxale : il a les caractéristiques phonologiques d'un clitique (trait de période 1), mais sa cliticité ne s'exprime qu'à mesure que le domaine s'élargit, d'abord marginalement et en avance sur l'élargissement général du domaine (du fait de la transition d'un état *pro drop* à un état non *pro drop* de la langue), puis de la même manière que pour les polysyllabes.

Si l'on admet un domaine donné, le fonctionnement de l'élision est exactement le même qu'en période 1 : les voyelles finales atones, flottantes, s'élident devant voyelle sous l'effet de *VCVIDES mais se maintiennent devant consonne et en fin de domaine sous l'effet de AssocV. Ces deux mécanismes sont donc suffisants pour rendre compte de l'élision, et la variation observée, partiellement diachronique (le domaine évolue progressivement), et partiellement synchronique (il y a une dimension stochastique), est due à la transcription des domaines à partir des constituants syntaxiques, et non à la phonologie elle-même.

14.4 Conclusion

Ce chapitre était dédié à l'analyse phonologique des données dont nous disposons, c'est-à-dire de celles issues de nos enquêtes informatiques sur le NCA et la métrique de textes de la BFM, ainsi que des témoignages que nous laissent la musication et les grammairiens du XVI^e s. Après un rapide résumé des données (sec. 14.1.1), ce chapitre s'est divisé en deux parties : la première traite des conditions proprement phonologiques de l'élision (sec. 14.2), et la seconde traite des domaines de l'élision (sec. 14.3).

Dans la section consacrée aux conditions phonologiques de l'élision, nous avons proposé de considérer que l'élision ne pouvait avoir lieu qu'en présence d'une voyelle flottante, et avons divisé les cas de flottement en

deux catégories : d'une part le flottement diacritique (sec. 14.2.1), qui distingue, en période 1, les monosyllabes fonctionnels qui s'élident de ceux qui ne s'élident pas, et d'autre part le flottement non diacritique (sec. 14.2.2), qui semble atteindre, à terme, tous les schwas de la langue. Nous avons avancé que le flottement diacritique était hérité d'un processus qui s'était appliqué en proto-a.fr., COND_{CV} (sec. 14.2.1.2). Ce processus rend flottante la voyelle finale des mots de structure CV, et uniquement de ceux-ci. Le flottement non diacritique, lui, est dû à une règle plus générale, qui semble cibler l'identité mélodique du schwa : COND._ə rend flottant les schwas, probablement d'abord en position finale, puis partout (sec. 14.2.2). Cette règle, contrairement à COND_{CV}, est donc encore active en synchronie en a.fr., et préfigure probablement le mutisme généralisé qui affectera le schwa en français moderne.

Ces conditions étant posées, nous avons proposé un mécanisme assez simple pour générer l'élision, que ce soit celle de la période 1 ou celle de la période 2 (sec. 14.2.3). Au sein d'un domaine computationnel donné, toute voyelle finale flottante s'élide devant initiale vocalique par l'effet de *VC_{VIDE}, qui élimine la position V à laquelle elle aurait pu être associée. En revanche, en fin de domaine et devant initiale consonantique, il n'y a pas de VC vide, la position V synchrone de la voyelle flottante n'est pas éliminée et AssocV associe donc la mélodie flottante avec cette position (sec. 14.2.3.4). Enfin, nous avons rapidement développé le fait que cette analyse permettait de prendre également en charge la réduction des hiatus internes, mais que sa motivation interne n'était pour autant pas réductible à un évitement du hiatus, avant de proposer une note sur la spécificité de la conjonction *que* (sec. 14.2.3.5).

Dans la section consacrée aux domaines de l'élision (sec. 14.3), nous avons commencé par définir que l'élision de la période 1 correspondait uniquement à celle des clitiques (sec. 14.3.1). Ceux-ci sont définis non seulement par leur position dans la structure syntaxique (sec. 14.3.1.1), mais également par un critère de taille : la période 1 de l'a.fr. n'admet que des clitiques monosyllabiques ; les polysyllabes, quels qu'ils soient, ne peuvent être clitiques (14.3.1.2). Nous nous sommes ensuite intéressés au domaine de l'élision en période 2 (sec. 14.3.2), et avons constaté que la phonologie prosodique était trop permissive pour permettre des déductions sur nos données (sec. 14.3.2.1). Néanmoins, l'examen des quelques occurrences d'adjectifs antéposés élidés et surtout des conditions syntaxiques de l'élision de *jo* nous ont permis de documenter quelques points de l'élargissement du domaine computationnel entre la période 1 et la période 2 (sec. 14.3.2.2). Enfin, nous avons déployé le même mécanisme d'élision que pour la période 1, et avons démontré que ce mécanisme parvenait aux résultats attendus, ce qui nous a amené à conclure que la seule différence entre l'élision de la période 1 et celle de la période 2 tient à leur domaine : en période 1, seuls les clitiques à voyelles finale flottante s'élident, tandis qu'en période 2 toute voyelle finale flottante (qui n'est pas en fin de domaine) s'élide.

L'analyse proposée ici peut donc se résumer à deux règles générant le flottement des voyelles finales atones : COND_{CV} et COND._ə. La première est active uniquement en proto-a.fr. et le flottement qu'elle génère est hérité par l'a.fr. ; la seconde est toujours active en synchronie en a.fr. L'élision en elle-même est obtenue par l'élimination des VC vides, tandis que la réalisation de la voyelle devant initiale consonantique et en position

finale de domaine est obtenue par AssocV. Enfin, les différences entre l'élision de la période 1 et celles de la période 2 ont pu être réduites à une seule caractéristique : en période 1, seuls les clitiques sont élidés, tandis qu'en période 2 cette restriction n'a plus lieu. À cela s'ajoute la découverte du fait que, selon nos données, la période 1 de l'a.fr. ne considère comme clitiques que les monosyllabes, et refuse ce statut aux polysyllabes fonctionnels.

15.1 Résumé des résultats et analyses

Dans cette thèse, nous nous sommes intéressés à l'élision en a.fr. Face au mutisme relatif des grammaires, nous avons voulu documenter précisément la situation de l'élision dans les textes médiévaux. Pour ce faire, nous avons développé un programme R (Comparalem : sec. 3.2) et l'avons exécuté sur le NCA. Nous avons ainsi pu étudier les alternances <e>/∅ des noms (chap. 4), des adjectifs féminins singuliers (chap. 5), des prépositions (chap. 6) et des adverbes (chap. 7), ainsi que les alternances <e,V>/∅ des articles (8), des pronoms personnels (chap. 9) et de certaines conjonctions (chap. 10). Une comparaison des taux de défauts devant initiale consonantique et vocalique nous a permis d'établir que, pour les adverbes, le pronom sujet *ele* et la conjonction *come*, le processus d'élision ne saurait être le coupable du défaut de <e>, puisque le défaut semble indépendant de l'initiale du mot suivant. Par ailleurs, Comparalem n'a repéré aucun défaut pour les noms, l'article *li* pl et le pronom *tu* CS. Quant aux autres items, leur élision peut être être systématique ou s'approcher de la systémativité (*la, la, te, de*), être optionnelle et dominante (*que*) ou optionnelle et minoritaire (*li* sg). Enfin, pour certains items comme le pronom sujet *jo*, l'élision graphique est variable dans le temps. Pour ce pronom, l'élision, inexistence au début de l'a.fr., devient la règle à la fin du Moyen Âge. Comme proposé en chap. 12, l'ensemble de ces données a été résumé à la généralisation suivante : ne sont élidés, au niveau graphique, que les mots qui avaient une structure CV atone en proto-a.fr.

Un second ensemble de données s'oppose à ces données graphiques : le témoignage de la métrique et de la musication, et celui des grammairiens du xvi^e s. (chap. 13). Contrairement à l'élision graphique, celle dont témoignent ces sources n'est plus restreinte aux seuls mots qui avaient une structure CV en proto-a.fr. : au sein d'un domaine qui peut contenir plusieurs clitiques ainsi que plusieurs mots non clitiques (fonctionnels ou lexicaux), toute voyelle finale atone est susceptible de s'élider. Nous avons proposé d'analyser cette tension entre l'élision graphique et l'élision métrique en envisageant que chacun témoignait d'un état de langue différent. Ainsi, l'élision graphique correspond assez bien à ce qu'on attendrait de l'élision dans la première période de l'a.fr., avant la fin du xii^e s., tandis que l'élision métrique correspond plutôt à ce que l'on attendrait de la seconde période de l'a.fr., lorsque le domaine computationnel s'élargit. Le fait que, dès *Alexis* et *Roland*, l'élision métrique soit déjà en place ne doit pas être considéré comme contredisant la diachronie que nous avons proposée. En effet, la métrique française ne connaît à l'époque pas de domaine computationnel plus petit que l'hémistiche, et ne peut donc pas proposer de domaine computationnel analogue à celui de la langue non versifiée.

Nous avons finalement pu proposer une analyse phonologique de l'ensemble des faits décrits. Nous proposons de considérer deux processus

phonologiques qui mènent au flottement d'une voyelle finale atone : d'une part COND_{CV}, qui a rendu flottante la voyelle des monosyllabes fonctionnels de structure CV en proto-a.fr. puis a cessé de s'appliquer, et d'autre part COND_ə, toujours active en a.fr., qui a pour effet de rendre flottant tout schwa final. Ce flottement mélodique, combiné à un mécanisme associant les mélodies flottantes lorsque des positions sont disponibles (AssosV) et à un mécanisme agissant sur les positions gabaritiques (*VC_{VIDES}), génèrent tous les cas d'élision observés. Dans cette analyse, la différence entre les deux périodes de l'élision ne tient donc qu'au domaine : en période 1, seuls les clitiques sont susceptibles de s'élider, alors que la cliticité n'est plus une condition nécessaire en période 2. Selon cette proposition, l'a.fr. ne considère comme clitiques que les monosyllabes fonctionnels, et refuse toujours ce statut aux polysyllabes, qu'ils soient lexicaux ou fonctionnels.

15.2 Apports

Au niveau théorique, l'apport principal de notre analyse tient au fait qu'il n'est nul besoin de faire appel à un processus spécialement dédié à la résolution des hiatus pour expliquer l'élision.

15.2.1 Apport théorique : la réduction des hiatus, conséquence et non cause

Pour ce qui est des deux périodes, nous avons montré que l'élision pouvait être réduite à un processus mécanique (prédictible, dont la variation est marginale). Ce processus, contrairement à l'analyse désormais classique de TRANEL (2000), ne tient pas nécessairement d'une logique anti-hiatique : les hiatus internes ne sont pas réduits au même rythme, ni de la même manière, que les séquences *schwa final - initiale vocalique*. De fait, si la résolution des hiatus était la cause de l'élision, on s'attendrait à ce qu'elle soit plus forte à l'intérieur du mot qu'entre les mots ; ce n'est pas le cas, puisque l'élision des clitiques est attestée avant, et de manière plus systématique, que la réduction des hiatus internes.

Nous avons proposé en sec. 14.2.3.4 que le schwa des hiatus internes était, au début de l'ancien français, associé à sa mélodie, et que la dissociation entre le schwa hiatique et sa mélodie n'était que lente et progressive, tant dans l'intensité de la réduction (pour chaque lemme) que dans sa diffusion lexicale (lemme par lemme). Dans le cas de l'élision, en revanche, le flottement de la mélodie vocalique finale précède l'ancien français ; il n'y a pas de moment dans l'histoire de l'a.fr. que l'on puisse identifier comme étant le point de dissociation entre mélodie et position. Si l'on considère, avec GUSSMAN et KAYE (1993), que l'élimination des VC vides peut être un outil universel de la phonologie, cet outil est paramétré comme actif entre deux mots dès les *Serments de Strasbourg*. Enfin, *li pl* et *tu CS* devant initiale vocalique ne sont jamais élidés, ce qui serait contrevenant à une contrainte anti-hiatique générale.

Ce point n'est pas sans conséquences théoriques. Il signifie, comme le proposait déjà MORIN (2005), que les phénomènes de sandhi externe ne sont pas réductibles à la tension qui serait inhérente aux hiatus.

TRANEL (2000). « Aspects de la phonologie du français et la théorie de l'optimalité ».

MORIN (2005). « La liaison relève-t-elle d'une tendance à éviter les hiatus? Réflexions sur son évolution historique ».

Selon la brève généalogie de l'étude de la liaison consonantique que cet auteur dresse, celle-ci est intimement liée aux canons de la récitation des vers, et norme et sur-norme, phonologie inconsciente et considérations esthétiques s'y mélangent. Or, dans les faits, un locuteur n'a aucune difficulté à réaliser une quantité conséquente de hiatus en français. Il convient donc de relativiser la toute-puissance d'une éventuelle contrainte anti-hiatique. L'analyse que nous proposons est tout à fait indépendante d'une putative mécanique anti-hiatique ; elle agit sur le gabarit, à partir de la proposition de GUSSMAN et KAYE (1993), et a pour effet (et non pour objectif) d'éliminer un schwa hiatique.

15.2.2 Apports en phonétique historique

Au-delà de la question traitée, nous pensons que cette thèse comporte également des éléments pertinents pour la phonétique historique du français.

15.2.2.1 L'élision et l'élargissement du domaine computationnel

Cette thèse permet d'inscrire l'élision au sein du processus d'élargissement du domaine computationnel, qui a lieu dans la seconde moitié du XII^e s. et la première moitié du XIII^e s. À notre connaissance, ce lien n'avait pas encore été proposé. Malgré le mutisme des données graphiques, l'interpolation nécessaire entre la situation de la période 1 et celle du XVI^e s., dont témoigne par anticipation la parophonologie métrique et musicale, permet d'établir que la causalité commune entre l'élargissement du domaine accentuel, la disparition de l'enclise et l'instauration de la liaison consonantique est aussi partagée par l'élision. De fait, la transition d'un système où l'élision dépend de la cliticité à un système où l'élision se libère de cette condition est le témoin fiable de cette intégration, somme toute assez naturelle, de l'élision au sein du paradigme des phénomènes de sandhi et d'accentuation. En retour, cette intégration peut renforcer — s'il en était besoin — les études portant sur l'élargissement du domaine computationnel en a.fr. Le petit développement consacré au pronom sujet *jo*, *je* (sec. 14.3.2.2) permet aussi de comprendre que la représentation phonologique de ce pronom n'est pas en cause dans ses occurrences non éliminées ; seul l'élargissement du domaine préside à son élision. En retour, une nouvelle fois, l'étude détaillée de la progression des domaines d'élision de *jo*, *je* de MARCHELLO-NIZIA (2015) permet de documenter assez précisément l'élargissement du domaine computationnel contenant un verbe.

15.2.2.2 Flottement et clitiques

Cette thèse permet également de comprendre l'idiosyncrasie apparente de *li pl* et *tu CS*, qui se refusent à l'élision. Pour ces deux formes, nous avons reconstruit une forme préhistorique dont la structure excède un couple CV : /lij/ et /tuw/. Le rejet de la cliticité de ces formes CVC ne leur est pas spécifiquement dédié : il concerne toutes les formes plus grandes qu'un couple de positions CV. De fait, *li pl* et *tu CS* s'intègrent naturellement au paradigme de tous les mots fonctionnels polysyllabiques qui, en période

1, ne connaissent pas l'élision (sauf cas de collocation grammaticalisée). Par la suite, en période 2, ces mots continuent à échapper à l'élision : seuls les polysyllabes profitent de l'élargissement du domaine pour s'élider, puisque leur schwa final est flottant. Rien de tel pour *li* et *tu*, qui héritent d'une voyelle finale bel et bien associée à sa position vocalique.

À cela s'ajoute la découverte du fait que seuls les monosyllabes peuvent être considérés clitiques en période 1 ; en effet, *ele*, *une*, *come* ne sont pas sujets à l'élision, et la seule chose qui les différencie des mots atteints par l'élision graphique, c'est le nombre de leurs syllabes. Il semble donc que l'état de langue dont témoigne la graphie n'accepte, en tant que clitiques, que des monosyllabes.

15.2.3 Apports techniques

Dans leur introduction au numéro de *Diachroniques* dédié à la phonétique historique, SÉGÉRAL et SCHEER (2015) font le constat que la phonétique historique du français est une *discipline sinistrée*. Parmi les raisons de ce sinistre qu'identifient les auteurs figure l'idée que la discipline n'a pas su intégrer les nouvelles méthodes de recherche, dont leur dimension informatique. Si nous ne sommes pas les premiers à développer une approche informatique à des fins de phonologie historique (cf. MARCHELLO-NIZIA 2015 ; RAINSFORD 2011b, 2020),¹ nous nous inscrivons dans une démarche récente, sinon naissante, pour ce qui est de l'a.fr. en tout cas. Nous avons développé deux outils pour interroger les sources médiévales à des fins d'analyse grapho-phonologique : Comparalem, et le PAM.

Si, dans cette thèse, Comparalem a été utilisé uniquement pour détecter des alternances <e>/Ø et <V>/Ø, le programme est en réalité capable d'analyser n'importe quelle variation linguistique qui serait réductible à deux variantes. La seule condition est d'arriver à fournir les *inputs* dont Comparalem a besoin ; du moment qu'un utilisateur est en mesure de lui fournir des tables de formes (avec leurs propriétés pertinentes) pour chaque variante, Comparalem peut comparer ces variantes. On peut donc envisager d'utiliser cet outil pour analyser n'importe quelle variation graphique ou grapho-phonologique de l'ancien français.

Le PAM, dont l'utilisation a été plus discrète dans ce document, permet également des analyses grapho-phonologiques, comme les remarques que nous avons pu faire sur les élisions métriques non transcrites dans *Alexis* et *Roland* (PREMAT et POGGIO [à paraître]). Il permet notamment de systématiser une démarche comme celle que MARCHELLO-NIZIA (2015) a menée par la vérification manuelle du statut métrique des pronoms étudiés.

Pour ces deux programmes, le développement va se poursuivre, et nous espérons être bientôt en capacité de les diffuser à la communauté, sous une forme plus simple à utiliser que l'architecture actuelle.

15.3 Ouvertures pour parfaire ce travail

Nous n'avons pas été en mesure, principalement par manque de temps et/ou à cause de l'hétérogénéité de certaines métadonnées, d'interroger

1: L'objectif de DEES (1980, 1985, 1987) étant essentiellement dialectologique et graphématique, nous ne considérons pas qu'il se soit inscrit dans une démarche de phonétique historique. Sa démarche de corpus et de cartographie, en revanche, sont des précédents précieux pour nous.

l'ensemble des facteurs qui peuvent influencer sur les variations <e>/Ø. Si l'examen systématique de la diatopie et de la diachronie ont permis d'isoler certains cas remarquables (p.ex. *ele* dont l'élision et l'apocope sont bien plus marquées au Sud-Ouest et demeurent raisonnables ailleurs), il conviendrait, pour une démarche plus exhaustive, d'interroger également le genre des textes, et surtout leur statut versifié ou non. On pourrait aussi envisager d'interroger, dans la diachronie, l'effet qu'aurait l'utilisation de la date de copie au lieu de la date de composition, ou l'effet que pourrait avoir l'écart entre la date de composition et la date de copie. De même, en diatopie, il ne serait pas intéressant de comparer nos données fondées sur les localisations de Dees et, uniquement lorsque les localisations de Dees sont absentes, sur les localisations philologiques, avec une approche où la localisation par défaut serait la localisation de composition, ou celle de copie.

Pour tous ces facteurs, l'utilisation de techniques statistiques plus avancées comme les modèles mixtes permettrait d'obtenir une vision du poids de chaque variable de manière plus satisfaisante que la comparaison manuelle de l'effet des variables, telle que nous l'avons faite. Toujours au niveau technique, si le code de Comparalem à évolué à mesure que notre analyse et nos besoins évoluaient, nous percevons désormais des défauts dans certains choix qui ont présidé à sa conception. Si ces défauts n'invalident aucune des mesures que nous avons proposées, il serait opportun de ré-examiner ces données lorsque le code du programme aura évolué. Par ailleurs, l'une des limites de nos analyses avec Comparalem tient à son corpus, le NCA, qui ne contient pas — ou très peu — de ponctuation. De fait, travailler sur un corpus ponctué ou, mieux encore, sur un corpus annoté en syntaxe comme le SRCMF (PRÉVOST ET STEIN 2013), permettrait non seulement d'isoler automatiquement les cas où un mot se trouve en fin de groupe syntaxique, mais également de documenter automatiquement, et de manière quantitative, les contextes syntaxiques dans lesquels se manifestent les quelques traces de l'élargissement du domaine.

PRÉVOST et al. (2013). *Syntactic Reference Corpus of Medieval French (SRCMF)*.

Enfin, si l'on voulait aller au bout du raisonnement de cette thèse, il conviendrait de compléter les données par trois études. La première consisterait à étudier la situation de monosyllabes fonctionnels que nous avons dû laisser de côté. Ainsi, à partir de la liste des mots sujets à élision en fr.mod., il conviendrait d'interroger l'élision de *si/se*, *ni/ne* et *ce*. Cela permettrait d'établir si l'élision du français moderne n'est, en somme, qu'une relique de l'élision de la première période de l'a.fr. C'est en partie la complexité du maniement des annotations du corpus, et en partie le temps dont nous disposions, qui ne nous ont pas permis de compléter cette exploration. La deuxième étude complémentaire porterait sur les cas où ce n'est pas la proclise mais l'enclise qui est à l'œuvre, dans les formes canoniques (*de + le* → *del*, *a + le* → *al*, etc.) et dans les cas de postposition du clitique (fr.mod. *dis-le*, *dit-il*, etc.). Une troisième étude complémentaire pourrait reprendre l'examen des données avec Comparalem, mais sur la version 4 du NCA, de manière à bénéficier de la lemmatisation de RRN Tagger, dont la qualité est meilleure que celle de Tree Tagger. Il y a fort à parier que Comparalem détecterait des variations à bas bruit en plus grand nombre avec cette lemmatisation plus fine, en particulier là où Tree Tagger a renoncé à attribuer un lemme à des formes non canoniques et rares.

ANNEXES

A

Abréviations, symboles, notations formelles

A.1 Abréviations et symboles

A.1.1 Périodisation

Sont ici regroupées les abréviations servant à la périodisation des langues.

lat. Latin.	t.a.fr. Très ancien français.	fr.p.class. Français pré-classique.
lat.class. Latin classique.	a.fr. Ancien français.	fr.class. Français classique.
lat.pop. Latin populaire.	m.fr. Moyen français.	fr.mod. Français moderne.
lat.tard. Latin tardif.	a.m.fr. Ancien et moyen français.	fr.cont. Français contemporain.
lat.méd. Latin médiéval.		

A.1.2 Abréviations générales

Sont ici regroupées les abréviations apparaissant régulièrement en plein texte.

& Soit abréviation du mot <i>et</i> , soit opérateur de jonction entre deux éléments.	(AND ² <i>Online Edition</i>) (2022).	CR_n Dans le corpus du NCA, le nombre <i>n</i> précédé de CR désigne la qualité d'une édition critique (cf. p. 22)
Δ Opérateur de différence entre deux grandeurs.	Anon. Anonyme.	DEAF <i>Dictionnaire Étymologique de l'Ancien Français (DEAF)</i> (1974-). La version citée est toujours celle du DEAFél.
× Soit position gabaritique, soit opérateur de multiplication.	BFM Base de Français Médiéval, cf. GUILLOT-BARBANCE, HEIDEN et LAVRENTIEV (2017).	DEAFél <i>Dictionnaire Étymologique de l'Ancien Français, version électronique (DEAFél)</i> (2010-). Cf. DEAF, DEAFpre.
σ Soit syllabe, soit écart type.	BNF Bibliothèque Nationale de France.	DEAFpré Sous ensemble du DEAFél, articles en version préliminaire. Avertissement : https://deaf-server.adw.uni-heidelberg.de .
Σ Somme.	C Consonne, ou position consonantique.	DECT <i>Dictionnaire Électronique de Chrétien de Troyes (DÉCT)</i> (2015).
~ La tilde note une opposition entre deux formes (cf. sec. A.2) ou l'approximation sur une valeur statistique.	C.C Groupe consonantique hétérosyllabique (cf. .CC).	dép. Dépendant(e) : décrit un test de Student à variables dépendantes, ou une mesure issue de ce test. Cf. indép.
*CC/_# Groupe consonantique illicite en position finale.	ca <i>Circa</i> , aux environs de la date mentionnée.	DÉRom <i>Dictionnaire Étymologique</i>
.CC Groupe consonantique tautosyllabique (cf. OL, TR).	cf. <i>Confer</i> , se reporter à.	
/Ø/ Objet phonologique sous-spécifié.	chap. Chapitre.	
∅ Absence d'objet, le plus souvent de réalisation phonétique.	col. Colonne.	
ALF <i>Atlas linguistique de la France</i> (GILLIÉRON et EDMONT 1902-1910).	compo. Composition. Précède ou suit l'indication d'un lieu ou d'une date de composition, par opposition à <i>ms.</i> dans cet usage.	
AND <i>Anglo-Norman Dictionary</i>	contra À la différence de, pour opposer un discours à un élément de bibliographie.	
	corr. Correction.	

- Roman (DÉRom)* (2022).
- dés.** Désinence.
- DMF** *Dictionnaire du Moyen Français (DMF)* (2020).
- épen.** Épenthèse, épenthétique.
- F** Fréquence.
- FEW** *Französisches Etymologisches Wörterbuch (FEW) informatisé* (WARTBURG 2022 [1922-1967]).
- fig.** Figure.
- fol.** Folio. Numéro de feuille dans une source; *r* et *v* après le numéro de folio indiquent que la référence s'applique au recto ou au verso du folio.
- GD** *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle* (GODEFROY 1891-1902).
- GDc** *Complément au Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle* (GODEFROY 1895-1902).
- germ.** Germanique (ancien germanique).
- GGHF** *Grande Grammaire Historique du Français* (MARCELLO-NIZIA, COMBETTES et al. 2020).
- GI** En CVCV, relation de gouvernement infrasegmental.
- Gvt** En CVCV, relation de gouvernement.
- h(.)** Hémistiche.
- indép.** Indépendant(e) : décrit un test de Student à variables indépendantes, ou une mesure issue de ce test. Cf. dép.
- L** Liquide.
- l.** Ligne.
- Lic.** En CVCV, relation de licenciement.
- litt.** Littérale (se dit d'une traduction).
- locus** Lieu, qu'il soit d'articulation ou de variation.
- m.** Moitié.
- MGH** *Monumenta Germaniae Historica*, coll. citée par ouvrages dans la bibliographique, cf. BORETIUS et KRAVSE (1887).
- MSn** Dans le corpus du NCA, le nombre *n* précédé de MS désigne la qualité d'une édition diplomatique (cf. p. 22).
- ms.** Manuscrit. Est soit une abréviation du mot *manuscrit*, soit précède ou suit l'indication d'un lieu ou d'une date de transcription, par opposition à *compo.* dans cet usage.
- N** Consonne nasale (*locus* non spécifié).
- N/A** Non Applicable ou (valeur) Non Attribuée. Le paramètre en question n'a pas de valeur attribuée.
- NCA** Nouveau Corpus d'Amsterdam, cf. STEIN, KUNSTMANN et GLEZGEN (2006).
- NIL ou nil** Absence de valeur pour une variable. Du latin *nil*, « rien, zéro ».
- O** Obstruante.
- OL** Groupe *obstruante + liquide* (cf. TR).
- p.** Page.
- PAM** Programme d'Analyse Métrique, cf. POGGIO et PREMAT (2019) et PREMAT et POGGIO ([à paraître]).
- pond.** Pondéré(e) : décrit une mesure portant sur le jeu de données dont les taux sont pondérés par le nombre d'occurrences des textes.
- q.** Quart.
- REW** *Romanisches Etymologisches Wörterbuch* (MEYER-LÜBKE 1911-1920).
- s.** Siècle.
- s. d.** Sans date.
- sec.** Section. Dans les renvois internes, *sec.* s'emploie sans distinction pour les sections, sous-sections et sous-sous-sections.
- t.** Tiers.
- TEI** *Text Encoding Initiative*. Consortium de standardisation des pratiques de philologie numérique.
- term.** Terminaison (de vers).
- TL** *Altfranzösisches Wörterbuch*, version électronique (TOBLER et LOMMATZSCH [2002] 1925-1976).
- TLFi** *TLFi* (1994).
- TR** Attaque branchante (cf. OL).
- TXM** Plateforme logicielle de textométrie, cf. HEIDEN, MAGUÉ et PINCEMIN (2010).
- V** Voyelle, ou position vocalique.
- v(.)** Vers.
- XML** *Extensible Markup Language*. Langage de balisage extensible, utilisé pour les éditions numériques de texte médiévaux.

A.1.3 Abréviations d'étiquettes morphosyntaxiques

Sont ici regroupées les abréviations composant les étiquettes morphosyntaxiques. Elles sont en bonne partie inspirées du jeu d'étiquettes CATTEX2009 (PRÉVOST, GUILLOT et al. 2013) utilisé par la BFM et par le jeu d'étiquettes utilisé par le NCA, mais nous les utilisons également hors des usages informatiques. Selon que l'usage est informatique ou non, la présence d'accents sur les voyelles varie, et est sans incidence. La casse est sans importance, mais les étiquettes composites sont décomposées par des points en minuscules et par des capitales lorsque l'étiquette est en petites capitales.

- CR** Cas Régime.
- CS** Cas Sujet.
- dir.** Direct (cas régime direct).
- fém.** Féminin.
- indir.** Indirect (cas régime indirect).
- masc.** Masculin.
- neu.** Neutre.
- pl(.)** Pluriel.

réfl. Réfléchi (pronom).	DÉTIND Déterminant indéfini.	VERCON Verbe conjugué.
sg(.) Singulier.	DÉTINT Déterminant interrogatif.	VERINF Verbe à l’infinitif.
ADJ Adjectif	DÉTPOS Déterminant possessif.	VERPPA Verbe au participe présent actif.
ADJDÉM Adjectif démonstratif.	DÉTREL Déterminant relatif.	VERPPP Verbe au participe passé passif.
ADJIND Adjectif indéfini.	NOM Nom (substantif), ou nom-natif.	COND Conditionnel.
ADJPOS Adjectif possessif.	NOMCOM Nom commun.	IMP Impératif.
ADJQUA Adjectif qualificatif.	NOMPRO Nom propre.	IND Indicatif.
ADJSUB Adjectif substantivé.	PRÉ Préposition.	SUBJ Subjonctif.
ADV Adverbe.	PRO Pronom.	FUT Futur.
CON Conjonction.	PRODÉM Pronom démonstratif.	IMP Imparfait.
CONCOO Conjonction de coordination.	PROIND Pronom indéfini.	PAR Parfait (passé simple).
CONSUB Conjonction de subordination.	PROPER Pronom personnel.	PQPAR Plus-que-parfait.
DÉT Déterminant	PROPOS Pronom possessif.	PRÉ Présent.
DÉTDEF Déterminant défini.	PROREL Pronom relatif.	
	VER Verbe.	

A.1.4 Dialectes

Nous utilisons des abréviations pour désigner les dialectes des textes. Celles-ci viennent en partie des méta-données de la BFM et du NCA. La méthode utilisée par chaque corpus pour assigner l’étiquette dialectale est différente; le NCA assigne, à côté des étiquettes de Dees, des étiquettes à la croisée de la géographie (non linguistique) et des dialectes. Lorsqu’une étiquette est inhabituelle, son origine est indiquée si elle ne provient que d’un seul corpus. Les noms de ville désignent une origine géographique précise du texte, et non l’aire dialectale générale rattachée à cette ville. Ainsi, l’étiquette *Liège* est plus précise que l’étiquette *liég.* (*liégeois*). Dans les notices du NCA, la localisation de DEES (1987) est suivie de l’origine dialectale ou géographique proposée par GLEÏGEN et GOUVERT (2007) et disponible dans les métadonnées du NCA (cf. sec. 2.1, p. 17 et sq., 21 et sq., 27 et sq. et 29 et sq.).

a.n. Anglo-normand.	lorr. Lorrain.	plus de précisions (NCA).
art. Dialecte de l’Artois.	Meuse Dialecte de la Meuse.	Soissonnais Dialecte de Soisson et de la région environnante.
Belgique Dialecte de Belgique, sans plus de précisions (NCA).	NE Dialecte du Nord-Est, sans plus de précisions (NCA).	tour. Dialecte tourangeau.
bourg. Bourguignon.	NO Dialecte du Nord-Ouest, sans plus de précisions.	wall. Wallon.
Centre Dialecte du Centre, sans plus de précisions (NCA).	Nord Dialecte du Nord, sans plus de précisions (NCA).	Beauvais Dialecte de la ville de Beauvais (NCA).
champ. Champenois.	norm. Normand.	La Rochelle Dialecte de la ville de La Rochelle (NCA).
Est Dialecte de l’Est, sans plus de précisions (NCA).	norm.-pic. Normanno-picard (NCA).	Liège Dialecte de la ville de Liège, dénomination plus précise que <i>liégeois</i> (NCA).
fla. Dialecte de la Flandre.	Oise Dialecte de l’Oise (NCA).	Lille Dialecte de la ville de Lille (NCA).
fr. cont. Français continental : n’importe quel dialecte à part l’anglo-normand.	orl. Orléanais.	Tournai Dialecte de la ville de Tournai (NCA).
fr.-it. « Franco-italien » (BFM).	Ouest Dialecte de l’Ouest, sans plus de précisions.	Troyes Dialecte de la ville de Troyes (NCA).
fr.-oc. « Franco-occitan » (BFM).	par. Parisien (BFM).	
fr.-pic. « Franco-picard » (BFM).	pic. Picard.	
franc. Francien (?) (un seul, NCA, Oise).	pic.-wall. Picardo-wallon (BFM).	
frc. Franc-comtois.	poit. Poitevin.	
hain. Dialecte du Hainaut.	saint. Saintonguais.	
liég. Liégeois.	SE Dialecte du Sud-Est, sans plus de précisions (NCA).	traits de l’Ouest Dialecte indéfini mais présentant principalement des traits occidentaux
	SO Dialecte du Sud-Ouest, sans	

(NCA).	mais présentant principale-	mérid. Méridional.
traits norm.? Dialecte indéfini	ment des traits occidentaux	occ. Occidental.
mais présentant principale-	et anglo-normands (NCA).	orient. Oriental.
ment des traits probable-	traits orl. Dialecte indéfini mais	sept. Septentrional.
ment normands (NCA).	présentant principalement	
traits occ. et a.n. Dialecte indéfini	des traits orléanais (NCA).	

A.2 Notations

Les caractères grecs ou en chasse fixe abc représentent des caractères quelconques.

< α >	Notation graphémique.	priété annotée d'un corpus numérique ; le plus
/ α /	Notation phonologique.	souvent l'identifiant d'un texte.
[α]	Notation phonétique.	
[α]	Un texte entre crochets dont le contenu est séparé des crochets par des espaces désigne un parenthésage de constituance et non une notation phonétique.	$/\alpha^\beta \alpha/$ Dans une notation phonologique, un caractère en exposant (ici β) renvoie à une mélodie flottante.
$\langle \alpha \rangle$	Notation de l'extramétricalité : la ou les syllabes entre chevrons ne sont pas associées à une position métrique.	$\alpha \sim \beta$ La tilde note l'opposition entre deux formes, pour les besoins de l'exposition (il ne s'agit pas nécessairement d'un contraste phonologique). [\sim devant une valeur numérique, cf. sec. A.1.2].
abc	Un texte en caractères à chasse fixe représente un fragment de code informatique ou une pro-	(α) α peut être présent ou absent.
		$\alpha : \beta$ Dans un exemple donné, α et β riment ensemble ou appartiennent à la même assonance.

B

Liste des généralisations, règles, principes, etc.

Adjectifs au féminin singulier (chap. 5)

- (5.57) **ADJFÉMSG : diffusion lexicale.**
Il n'y a pas de phénomène de diffusion lexicale du défaut de <e> observable. → p. 167
- (5.58) **ADJFÉMSG : fréquence.**
La corrélation entre taux de défaut de <e> et fréquence des lemmes est faible, non significative et inverse. Elle représente un effet de filtre quantitatif, et non un effet linguistique.
($\rho = -0,29$; $p = 0,44$) → p. 168
- (5.60) **ADJFÉMSG : collocations.**
Un défaut de <e> semble être encouragé par la collocation avec un nom à initiale vocalique. Cela ne concerne qu'une minorité de lemmes et des instances des collocations en question. → p. 170
- (5.61) **ADJFÉMSG : contexte.**
Le contexte prévoicalique est associé à plus de défauts de <e> que le contexte préconsonantique.
($t = 4,99$; $p < 0,005$)
L'effet n'est pas stable dans le temps et l'espace. → p. 171
- (5.62) **ADJFÉMSG : diachronie.**
On observe une concentration de taux de défaut de <e> plus élevés dans la seconde moitié du XI^es. Cet effet n'est pas homogène sur l'ensemble du territoire (5.63, 5.67). → p. 171
- (5.66) **ADJFÉMSG : spécificité de l'a.n.**
L'anglo-normand (a) se distingue fortement des dialectes continentaux par (b) son hétérogénéité, (c) ses nombreux textes à forts taux de défaut de <e> et (d) son importante susceptibilité à un effet de filtre selon le nombre de textes disponibles.
a) $_ \#V$: $t = 4,02$; $p < 0,005$;
b) $_ \#V$: $\sigma = 0,23$ vs a.fr. continental $\sigma = 0,08$;
c) $_ \#V$: moy. = 11% vs a.fr. continental 4%,
 $_ \#C$: moy. = 4% vs a.fr. continental 1%;
d) cf. (b), tables 5.4 et 5.5.
→ p. 176
- (5.67) **ADJFÉMSG : poids de l'a.n. dans les données.**
Étant donné (5.66) et le nombre élevé de textes anglo-normands (48), plus il y a de textes anglo-normands, plus les défauts de <e> sont nombreux. → p. 177
- (5.68) **ADJFÉMSG : poids de la variation.**
Rapportés à l'ensemble des ADJFÉMSG, les défauts de <e> sont extrêmement marginaux (a). Néanmoins, il semble qu'il s'agisse d'une *variation à bas bruit* plutôt que d'un simple *bruit de fond* (b).
a) 0,6% des occ. d'ADJFÉMSG,
2,4% des occ. des 45 lemmes;
b) Cf. (5.61, 5.62, 5.63, 5.66, 5.67).
→ p. 177

Prépositions (chap. 6)

- (6.79) **PRÉP : sélectivité lexicale.**
Les seuls lemmes en C_ə et TR_ə susceptibles de connaître le défaut de schwa le connaissent, mais des lemmes en C.C_ə pourraient le connaître et ne le connaissent pas. → p. 234
- (6.80) **PRÉP : sélectivité syntaxique.**
D'autres lemmes en CC_ə (mais pas en OL_ə) pourraient connaître le défaut de <e>, mais seuls « entre », « contre » et « encontre » le connaissent. Cela peut provenir de configuration syntaxique et/ou fréquentielles (cf. 6.83) propres à ces lemmes. → p. 234
- (6.81) **PRÉP : diffusion lexicale.**
Il n'y a pas de phénomène de diffusion lexicale du défaut de <e> observable. → p. 234
- (6.82) **PRÉP : fréquence.**
Il est inapproprié de chercher un effet de la fréquence d'emploi sur les taux de défaut de <e> de cinq lemmes. → p. 234
- (6.83) **PRÉP : collocations.**
Le défaut de <e> de « entre », « contre » et « encontre » semble être favorisé par la collocation avec le PRO_{PERS3pl} CR « eux » et, à plus faible fréquence, avec d'autres éléments. → p. 236
- (6.84) **PRÉP : contexte, « de », « entre ».**
Le contexte prévoicalique est le seul à être associé à des défauts de <e>, et cette association est majoritaire.
(« De » : $t = 86$; $p < 0,005$; moy. $_ \#V = 89\%$)
(« Entre » : $t = 23$; $p < 0,005$; moy. $_ \#V = 68\%$)
L'effet n'est pas stable dans le temps et l'espace. → p. 236
- (6.85) **PRÉP : contexte, « (en)contre ».**
Le contexte prévoicalique est le seul à être associé à des défauts de <e>, et cette association est minoritaire.
($t = 4,62$; $p < 0,005$; moy. $_ \#V = 24\%$)
La stabilité de l'effet ne peut être mesurée étant donnée la faible quantité des occurrences. → p. 236
- (6.86) **PRÉP : contexte, « (de)sor »**
Le contexte phonologique n'a pas d'effet sur les défauts de <e> des PRÉP « sor » et « desor », dont le <e> semble avoir majoritairement disparu.
($t = 0,02$; $p = 0,98$). → p. 237
- (6.87) **PRÉP : diachronie (a).**
Par élimination progressive des *outliers*, la préposition « de » fait augmenter le taux de défaut de <e> devant initiale vocalique. Il n'y a pas probablement pas d'évolution diachronique pertinente dans le contexte préconsonantique. → p. 238
- (6.88) **PRÉP : diachronie (b).**
Devant initiale vocalique, on observe un plateau ou un pic, selon que les taux ne sont pas ou sont pondérés par la quantité d'occurrences, dans la première moitié du XIII^e s. (cf. 6.93). → p. 238
- (6.89) **PRÉP : diatopie.**
Devant initiale consonantique, le domaine est très homogène, et les manquements temporaires à cette homogénéité semblent n'être dus qu'au hasard.
Devant initiale vocalique, le domaine continental est assez homogène, avec simplement un axe Centre - Sud-Est dont les taux sont légèrement moins forts dans la seconde partie des XI^e et XII^e s. L'Angleterre se détache clairement du domaine continental à partir de la seconde moitié du XII^e s. → p. 239

(6.92) **PRÉP : spécificité de l'a.n.**

« De » : l'a.n. (a) se distingue des dialectes continentaux par (b) son hétérogénéité et (c) ses nombreux textes à faibles taux de défaut de <e>.

- a) $_ \#V : t = -5.33; p < 0,005;$
 b) $_ \#V : \sigma = 0,32$ vs a.fr. continental $\sigma = 0,09;$
 c) $_ \#V : \text{moy.} = 68\%$ vs a.fr. continental 95%.

« (De)sor » : l'a.n. (a) est plus homogène que l'a.fr.cont., et (b) pratique systématiquement le défaut de <e> devant initiale vocalique.

- a) $_ \#V : \sigma = 0$ vs a.fr. continental $\sigma = 0,16;$
 $_ \#C : \sigma = 0,02$ vs a.fr. continental $\sigma = 0,1.$
 b) $_ \#V : \text{moy.} = 100\%$ vs a.fr. continental moy. = 96%.

→ p. 243

(6.93) **PRÉP : poids de l'a.n. dans les données.**

Étant donné (6.92 « de »), plus il y a de textes anglo-normands, moins les défauts de <e> devant initiale vocalique sont nombreux.

→ p. 244

(6.94) **PRÉP : poids de la variation.**

Rapportés à l'ensemble des PRÉP, les défauts de <e> sont globalement minoritaires (a), mais concernent tous les lemmes prépositionnels se terminant en $C\emptyset$ et $OL\emptyset$ (b). Leur statut varie fortement en fonction des lemmes et des contextes (c).

- a) Défauts : 6% des occ. de PRÉP, 16% des occ. des 6 lemmes.
 b) Les 6 lemmes retenus contiennent 100% des lemmes en $-C\emptyset$ et ~100% de ceux en $-OL\emptyset$
 c) « de » 0% $_ \#C$ 89% $_ \#V$
 « entre » 0% $_ \#C$ 68% $_ \#V$
 « (en)contre » 0% $_ \#C$ 24% $_ \#V$
 « (de)sor » 95% $_ \#C$ 95% $_ \#V$

→ p. 245

Adverbes (chap. 7)(7.31) **ADV : sélectivité lexicale.**

Nous estimons qu'à peu près la moitié des lemmes adverbiaux réunissant les conditions pour les défauts de <e> ont des occurrences en défaut de <e>.

→ p. 289

(7.32) **ADV : diffusion lexicale.**

L'apparition extrêmement précoce des premières occurrences défectueuses semble plaider pour une absence de diffusion du défaut au sein du corps lexical. En revanche, le léger retard de « ariere » et « derriere » et leurs taux inférieurs semblent témoigner, mais de manière peu consolidée, de la fin du processus de diffusion lexicale.

→ p. 289

(7.33) **ADV : fréquence.**

Les données plaident pour une absence d'effet de la fréquence lexicale sur les taux non contextuels de défaut de <e>. En revanche, la fréquence semble influencer sur la dimension stochastique du défaut de <e> (cf. 7.36).

→ p. 290

(7.34) **ADV : collocations.**

Dans *Roland*, la collocation grammaticalisée *ensembl od* est toujours en défaut de <e> et représente le seul contexte où le lemme « ensemble » est en défaut de <e>.

→ p. 291

(7.35) **ADV : contexte.**

Hormis pour « ensemble » (cf. 7.34), l'effet statistique faible du contexte subséquent sur la génération de défauts de <e> ne semble pas pertinent sur le plan de l'analyse linguistique.

($t = -2,5; p < 0,05; \text{moy. } _ \#C = 74\%; \text{moy. } _ \#V = 68\%$)

Ce faible effet est particulièrement instable dans le temps et dans l'espace.

→ p. 291

(7.36) **Adv : lemmes et contexte.**

La moitié des lemmes dont la fréquence est la moins élevée (a) montre une sensibilité erratique au contexte (b). Cette différence n'est probablement due qu'à la dimension stochastique du phénomène en jeu, qui ne s'exprime qu'en-deçà d'un certain seuil de fréquence (cf. 7.33).

- a) $F \leq 552$ vs $F \geq 936$ pour les autres lemmes,
 b) « onque », « mar » : $_ \#V$ favorise le défaut de <e>
 « derriere », « desor » : $_ \#C$ favorise le défaut de <e>.

→ p. 291

(7.37) **Adv : diachronie.**

Les taux de défaut de <e> sont constants. L'observation d'une tendance à la croissance est due à l'influence de l'anglo-normand (cf. 7.39).

→ p. 294

(7.38) **Adv : diatopie.**

Un arc de cercle Angleterre - Centre - Sud-Est du domaine comporte tous les taux inférieurs à la moyenne. Cet arc est plus large devant initiale vocalique. Les taux les plus élevés se situent à l'Est du domaine.

Cette distribution diatopique n'est stable qu'à partir du xm^e s.

→ p. 295

(7.39) **Adv : spécificité de l'a.n.**

L'a.n. (a) se distingue des dialectes continentaux par (b) ses nombreux textes à faibles taux de défaut de <e>, mais (c) son hétérogénéité n'est qu'à peine supérieure à celle du domaine continental.

- a) $_ \#V$: $t = -7,63$; $p < 0,005$;
 b) $_ \#V$: moy. = 31% vs a.fr. continental moy. = 73%;
 c) $_ \#V$: $\sigma = 0,31$ vs a.fr. continental $\sigma = 0,26$.

→ p. 297

(7.40) **Adv : poids de l'a.n. dans les données.**

Étant donné (7.39), plus il y a de textes anglo-normands, moins les défauts de <e> (devant initiale vocalique comme consonantique) sont nombreux (cf. 7.37)

→ p. 297

(7.41) **Adv : poids de la variation.**

Rapportés à l'ensemble des Adv, les défauts de <e> sont globalement minoritaires (a), mais sont majoritaires pour les lemmes concernés (b) et ceux-ci représentent une part conséquente des lemmes susceptibles de connaître un tel défaut (cf. 7.31).

- a) Défauts : 6% des occ. d'Adv,
 b) Défauts : 75% des occ. des 8 lemmes retenus.

→ p. 298

Articles (chap. 8)(8.36) **ART : sélectivité lexicale.**

Il est impossible, sur la base de critères phonologiques, de prédire quels articles sont atteints par l'élision.

→ p. 323

(8.37) **ART : sélectivité morpho-syntaxique.**

Ni la morphologie ni la syntaxe ne sont en mesure de prédire quels articles sont sujets à l'élisions.

→ p. 323

(8.38) **ART : restrictivité phonologique.**

Si la seule phonologie ne peut prédire quels seront articles atteints par l'élision, elle semble en revanche fournir un critère de différenciation de l'article indéfini féminin singulier, qui est le seul à être bisyllabique (*une*). Cet article se comporte comme les adjectifs féminins singuliers (cf. 8.39).

→ p. 324

(8.39) **ARTINDÉFFÉM sg : grammaticalisation.**

Environ la moitié des occurrences *un* pour *une* $_ \#V$ appartiennent à la construction *un(e) + autre + [nom]^{FÉM}*, qui semble donc se grammaticaliser.

→ p. 324

- (8.40) **ARTDÉF : contexte, le, la**
 Le contexte prévoicalique agit comme condition nécessaire et presque suffisante au déclenchement du défaut des articles définis masculins singuliers CR *le* et féminins singuliers CS-R *la, le*.
 (MASC CR sg : $t = 350$; $p < 0,005$; moy. $_\#V = 99\%$)
 (FÉM CS-R sg : $t = 532$; $p < 0,005$; moy. $_\#V = 99\%$) → p. 324
- (8.41) **ARTDÉF : contexte, li CSsg**
 Le contexte prévoicalique agit comme condition nécessaire mais non suffisante au déclenchement du défaut des articles définis masculins singuliers, dans la mesure où cette élision n'est pas systématique.
 ($t = 17$; $p < 0,005$; moy. $_\#V = 33\%$) → p. 324
- (8.42) **ARTINDÉF : contexte, une CS-Rsg**
 Le contexte prévoicalique agit comme condition favorable, mais ni nécessaire ni suffisante, au déclenchement du défaut des articles indéfinis féminins singuliers *une*.
 ($t = 6$; $p < 0,005$) → p. 325
- (8.43) **ART : diachronie et diatopie.**
 L'ARTDÉFMASC CSsg *li* s'élide plus à l'Ouest et au Sud-Ouest qu'ailleurs; l'ARTINDÉFFÉMCS-Rsg *une* perd son *-e* principalement en Angleterre et en Vendée, Deux-Sèvres. → p. 325

Pronoms personnels (chap. 9)

- (9.57) **PROPERS : sélectivité lexicale.**
 Il est impossible, sur la base de critères phonologiques, de prédire quels pronoms sont atteints par l'enclise et l'élision. → p. 382
- (9.58) **PROPERS : sélectivité morpho-syntaxique.**
 Ni la morphologie ni la syntaxe ne sont en mesure de prédire quels pronoms sont atteints par l'élision. → p. 382
- (9.59) **PROPERS : diffusion lexicale.**
 L'élision est présente d'emblée pour les pronoms régimes, et ne se développe que progressivement pour le pronom sujet de 1sg. → p. 382
- (9.60) **PROPERS : restrictivité phonologique.**
 Si la seule phonologie ne peut prédire quels seront les pronoms personnels sujets à l'élision, elle semble en revanche fournir un critère d'exclusion du pronom sujet féminin de 3sg, qui est le seul à être bisyllabique (*ele*). Ce pronom se comporte comme les adverbes. → p. 383
- (9.61) **PROPERS (enclise) : grammaticalisation.**
 L'enclise ne concerne réellement que le pronom régime de 3sg, probablement suite à un processus de grammaticalisation. La fréquence d'emploi des pronoms régimes de 3sg est compatible avec une telle grammaticalisation. → p. 383
- (9.62) **PROPERS : contexte, 1-2sg CR : me, te**
 Le contexte prévoicalique agit comme condition nécessaire et presque suffisante au déclenchement du défaut des pronoms régimes de 1 (*m(e)*) et 2sg (*t(e)*).
 (1sg CR : $t = 155$; $p < 0,005$; moy. $_\#V = 97\%$)
 (2sg CR : $t = 92$; $p < 0,005$; moy. $_\#V = 96\%$) → p. 384
- (9.64) **PROPERS : contexte, 1sg CS : je, jo**
 Le contexte prévoicalique agit comme condition nécessaire mais non suffisante au déclenchement du défaut des pronoms sujet de 1sg (*j(e) etc.*), ce qui est dû à l'absence de discrimination des formes atones et à l'extension seulement progressive des contextes syntaxiques de l'élision.
 ($t = 22$; $p < 0,005$; moy. $_\#V = 48\%$) → p. 384

(9.65) **PROPers : diachronie.**

L'élision et son évolution dépend du pronom impliqué (a-c), et l'enclise du pronom régime de 3sg est en régression (d). La variation du pronom sujet fém. 3sg est surtout diatopique (e).

- a) CR sg. __#V stable : 96-99%
- b) CS 1sg. __#V croissance de 0 à ~60%
- c) CS 2sg. __#{C, V} stable : 0%
- d) CR 3sg (MASC) __#C décroissance de ~40% à ~0%
- e) CS 3sg. FÉM __#{C, V} tendance à la décroissance

→ p. 385

(9.66) **PROPers : diatopie.**

L'élision du pronom sujet de 1sg *j(e)* se concentre à l'Ouest et sur un axe de la Normandie à la Meuse (a), les défauts du pronom sujet de 3sg FÉM *el(e)*, devant #V comme devant #C, se concentrent également à l'Ouest, tandis que l'enclise du pronom régime de 3sg se concentre au Nord-Est (c). L'anglo-normand n'a pas de comportement général spécifique.

- a) 1sg CS __#V Concentration à l'Ouest et de
la Normandie à la Meuse
- b) 3sg FÉM __#{C, V} Concentration à l'Ouest
- c) 3sg CR (MASC) __#C Concentration au Nord-Est

→ p. 385

Conjonctions : *Que*, *-que* et *come* (chap. 10)(10.24) **CONJ : sélectivité lexicale.**

Tous les lemmes conjonctifs en (-)Cə ou (-)CV, susceptibles de connaître le défaut de schwa, le connaissent. → p. 436

(10.25) **CONJ : sélectivité syntaxique.**

La conjonction « come » répond à une répartition syntaxique archaïque, qui se perd au centre du domaine et se répand ensuite depuis ce centre. → p. 436

(10.26) **CONJ : contexte, base + *-que* et élément *que*.**

Le contexte prévoicalique est le seul à être associé à des défauts de <e>, et cette association n'est pas systématique.

(base + *-que* : $t = 26$; $p < 0,005$; moy. __#V = 69%)

(élément *que* : $t = 32$; $p < 0,005$; moy. __#V = 41%)

L'effet n'est pas stable dans le temps et l'espace. → p. 437

(10.27) **CONJ : contexte, « comme ».**

Les défauts de <e> ne semblent pas sensibles au contexte phonologique.

($t = -2$; $p = 0,06$; moy. __#V = 62%; moy. __#C = 67%)

L'effet n'est pas stable dans le temps et l'espace. → p. 437

(10.28) **CONJ : diachronie.**

Il n'y a pas de réelle tendance diachronique pour l'élision des conjonctions en base + *-que* et de l'élément *que*. Pour « come », la tendance est à l'évacuation progressive du type *com* au profit du type *come*, dans tous les contextes syntaxiques. → p. 438

(10.29) **CONJ : diatopie.**

Les finales atones sont significativement plus maintenues sur une diagonale Nord-Ouest - Sud-Est qu'aux marges de celle-ci. → p. 438

(10.30) **CONJ : poids de la variation.**

Les alternances avec \emptyset concernent toutes les CON qui se terminent par un schwa (a) et tous les PROCON qui se terminent par <e> ou <i> (b). Les défauts sont fortement marqués (c). L'élément *que* et « come » ont une fréquence absolue élevée (d).

- a) CON : les 31 lemmes retenus contiennent 100% des lemmes en (-)C \emptyset ;
 b) PROCON : les 3 lemmes sujets à alternance <e,i>/ \emptyset concernent 100% des lemmes en -CV;
 c) *base* + -*que* 0% __#C 69% __#V
 élément que 0% __#C 41% __#V
 « come » 67% __#C 62% __#V
 d) élément *que* : F = 90 250 occ.
 « come » : F = 10 039 occ.

→ p. 439

Généralisations sur les données graphiques (chap. 12)

- (12.4) CONDITION MORPHOSYNTAXIQUE
 Ne sont candidats à l'élision que les mots fonctionnels. → p. 496; remplacé par (14.1).
 (12.5) CONDITION DE STRUCTURE CV (v1)
 Un mot fonctionnel n'est candidat à l'élision graphique que s'il se réduit à une structure *consonne* + *voyelle*. → p. 496; remplacé par (14.1).
 (14.1) CONDCV
 Un mot n'est candidat à l'élision graphique que s'il se réduit à une structure *consonne* + *voyelle* atone. → p. 542; remplacé par (14.10).
 (12.14) CONDCOLL (v1)
 Un mot peut perdre son <e> final devant un second terme à initiale vocalique avec lequel il forme une collocation entièrement grammaticalisée. → p. 503; remplacé par (12.24).
 (12.24) CONDCOLL (v2).
 Un mot peut perdre son <e> final devant un second terme à initiale vocalique avec lequel il forme une collocation, y compris si celle-ci n'est pas entièrement grammaticalisée. → p. 505

L'élision en métrique et musication (chap. 13)

- (13.6) 8s [× × × × × × × '×]_v → p. 512; remplacé par (13.9).
 (13.7) 10s [[× × × '×]_h [× × × × × '×]_h]_v → p. 512; remplacé par (13.10).
 (13.8) 12s [[× × × × × '×]_h [× × × × × '×]_h]_v → p. 512; remplacé par (13.11).
 (13.9) 8s × × × × × × × '× → p. 513
 (13.10) 10s × × × '× × × × × × '× → p. 513
 (13.11) 12s × × × × × '× × × × × × '× → p. 513
 (13.15) BIJEC.
 L'association entre syllabes et positions se fait selon un rapport bijectif. → p. 514
 (13.16) '× ⇒ 'σ
 Une position métrique proéminente est associée avec une syllabe accentuable. → p. 514
 (13.20) × ⇒ σ
 Toute position métrique non proéminente doit être associée à une (et une seule) syllabe. → p. 515
 (13.21) '× ⇒ 'σ
 Une position métrique proéminente est associée avec une syllabe accentuable. → p. 515

$$(13.22) \sigma \Rightarrow \times \quad \neg / \left[\begin{array}{c} \times \\ \updownarrow \\ \sigma \quad _ \end{array} \right]$$

Toute syllabe atone doit être associée à une (et une seule) position métrique, sauf (–) si elle est la posttonique d'un paroxyton dont la tonique est associée à une position proéminente (= sauf lorsqu'elle se trouve immédiatement après une syllabe accentuable associée à une position proéminente et qui appartient au même mot). → p. 515

$$(13.23) \sigma \Rightarrow \{ \times, \times \}$$

Toute syllabe accentuable doit être associée à une position, que celle-ci soit proéminente ou non. → p. 515

$$(13.24) \blacktriangledown \Rightarrow \sigma$$

Tout événement musical doit correspondre à une syllabe. → p. 516

$$(13.25) \sigma \Rightarrow \blacktriangledown$$

Toute syllabe doit correspondre à un événement musical. → p. 516

$$(13.33) \acute{E}LID(\sigma\sigma)$$

Toute voyelle finale atone de polysyllabe immédiatement suivie d'une initiale vocalique s'élide. → p. 522

$$(13.34) \acute{E}LID(\sigma)$$

Toute voyelle finale atone de monosyllabe immédiatement suivie d'une initiale vocalique peut s'élider. → p. 523

$$(13.45) \acute{E}LID.DOM.^{diac.}$$

L'élision métrique a pour domaine un constituant métrique. → p. 529

$$(13.46) \acute{E}LID.DOM.$$

$$\{ \emptyset, V \} \rightarrow \emptyset \neg / \left[\begin{array}{c} \times \\ \updownarrow \\ \sigma \quad _ \end{array} \right]$$

Une syllabe phonologiquement élidable subit l'élision métrique lorsqu'elle n'est pas (–) la posttonique d'un mot dont la tonique est associée avec une position proéminente. → p. 529

$$(13.48) PARAPH.\acute{E}LID.DOM.$$

La parophonologie force le domaine de l'élision, quel qu'il soit, à correspondre à l'un de ses constituants. → p. 533

Analyse phonologique (chap. 14)

$$(14.6) \acute{E}LID.FLOT.$$

Les voyelles flottantes sont candidates à l'élision, les voyelles non flottantes ne sont pas candidates à l'élision.. → p. 545; remplacé par (14.32).

$$(14.10) COND_{CV} (v3)$$

Toute voyelle finale atone d'un mot de structure CV devient flottante. → p. 545

$$(14.15) COND_{\emptyset\#}$$

Tout schwa final devient flottant. → p. 547; remplacé par (14.17).

$$(14.17) COND_{\emptyset}$$

Tout schwa est devient flottant. → p. 548

$$(14.28) ASSOCV$$

Lorsqu'une mélodie vocalique flottante peut, sans croiser de ligne d'association, s'associer avec une position V vide, elle le fait. → p. 554

$$(14.32) *VC_{VIDE}$$

Toute séquence de positions VC où V et C sont vides est éliminée. → p. 554

- (14.47) $\text{DOM} \Rightarrow \text{MOTLEX}$
Un domaine computationnel correspond à un mot lexical. → p. 562
- (14.48) $*\text{DOM} = \text{MOTFONC} [\text{MOTLEX}]$
Pour être traité par la phonologie, un mot doit appartenir à un domaine computationnel. → p. 562
- (14.49) $\text{DOM} = [\text{MOTFONC} \text{MOTLEX}]$
Un mot fonctionnel peut appartenir au domaine du mot lexical suivant lorsqu'il a un nœud commun avec celui-ci de niveau suffisamment bas dans l'arbre syntaxique. → p. 563; remplacé par (14.51).
- (14.50) $\text{DOM} = [\text{MOTFONC}]$
Lorsque le mot fonctionnel est le seul composant de son nœud syntaxique et que celui-ci est suffisamment indépendant du nœud syntaxique subséquent, le mot fonctionnel peut former son propre domaine. → p. 564
- (14.51) $\text{DOM} = [\text{MOTFONC}(\sigma) \text{MOTLEX}]$
Un mot fonctionnel peut appartenir au domaine du mot lexical suivant lorsqu'il a un nœud mère avec celui-ci de niveau suffisamment bas dans l'arbre syntaxique, et qu'il ne contient qu'une seule syllabe. → p. 565
- (14.52) $[\text{MOT}(\sigma\sigma\dots)] \Rightarrow \text{DOM}$
Lorsqu'un mot est polysyllabique, il doit correspondre à son propre domaine computationnel. → p. 565

Corpus informatiques : liste des textes

C.1 NCA : liste des textes

Lorsque l'indication de MS est absente, c'est que le sigle du MS au sein de la tradition (s'il existe), n'est pas donné dans les métadonnées du NCA. Lorsque la localisation de Dees est placée entre crochets, c'est que la variable `regionDees` est absente, mais que l'information de Dees est restituée à partir d'une autre variable (`codeRegional` ou `coefficientRegional`). Les informations de composition (*compo.*), de copie (*ms.*) et de qualité éditoriale (MS-1-3, CR-1-3) viennent du travail sur la bibliographie philologique de l'équipe du NCA; cf. 2.1.3, p. 20 *et sq.* pour la discussion de ces métadonnées. Lorsque l'édition est identifiée et publiée, elle est indiquée après l'évaluation de la qualité, soit par le nom de l'auteur soit par le nom de l'éditeur, selon que l'auteur est anonyme ou non, et selon le type de publication.

La liste donnée correspond à l'ensemble des textes du NCA. Les notices des textes cités en plein texte ont été reprises par nos soins (établissement du lien vers le DEAF, éventuelles corrections, et établissement de la notice de l'édition utilisée lorsque celle-ci est publiée). Les notices des textes non cités auparavant dans ce document, dont l'identifiant se termine par un astérisque (qui est de notre fait), ont été extraites automatiquement des métadonnées; le lien vers le DEAF n'est pas fonctionnel, et l'édition utilisée ne dispose pas de notice bibliographique autonome. Dans ces notices non vérifiées, les identifiants DEAF se terminant par un astérisque sont des propositions d'identifiant faites par l'équipe du NCA, et ne correspondent donc pas à un identifiant DEAF déjà existant; dans les notices vérifiées, ces identifiants fictifs ont été retirés.

- abe** : JEAN DE MEUN (1288*). *Traduction de la première épître de P. Abélard*. Dees : région parisienne; compo. frc., ca 1280; ms. Paris, ca 1395; MS3 (JEAN DE MEUN 1934).
- abreja** : JEAN PRIORAT DE BESANÇON (1290*). *Li abrejance de l'ordre de chevalerie, v. 1-2266*. Dees : [Franche-Comte]; compo. frcomt., ca 1290; ms. frcomt., ca 1300; MS (JEAN PRIORAT DE BESANÇON 1897); DEAF : JPrioratR.
- ailea** : RAOUL DE HOUDENC (1207*). *Le roman des ailes*. Ms. A; Dees : Yonne; compo. pic., ca 1207; ms. frc., ca 1288; MS (RAOUL DE HOUDENC 1983).
- aileb** : RAOUL DE HOUDENC (1207*). *Le roman des ailes*. Ms. B; Dees : Nièvre, Allier; compo. pic., ca 1207; ms. a.n., ca 1290; MS (RAOUL DE HOUDENC 1983).
- ailed** : RAOUL DE HOUDENC (1207*). *Le roman des ailes*. Ms. D; Dees : Aube; compo. pic., ca 1207; ms. frc., ca 1290; MS (RAOUL DE HOUDENC 1983).
- aileg** : RAOUL DE HOUDENC (1207*). *Le roman des ailes*. Ms. G; Dees : Angleterre; compo. pic., ca 1207; ms. a.n., ca 1300; MS (RAOUL DE HOUDENC 1983).
- aileo** : RAOUL DE HOUDENC (1207*). *Romanz des Eles*. MS O; Dees : Nord; compo. pic., ca 1207; ms. NO, ca 1310; MS (SUCHIER 1891); DEAF : ElesoS.
- ailet** : RAOUL DE HOUDENC (1207*). *Le roman des ailes*. Ms. T; Dees : Wallonie; compo. pic., ca 1207; ms. wall., ca 1300; MS (RAOUL DE HOUDENC 1983).
- aiol** : ANON. (1160*). *Aiol*. Dees : Somme, Pas-de-Calais; compo. pic., ca 1160; ms. pic., ca 1275; MS1 (NORMAND et RAYNAUD 1877); DEAF : Aiol^{1/2}N.
- alexo*** : ANON. (1210*). *The Oxford Version of the Vie de saint Alexis, ms. O*. Dees : [Wallonie]; compo. nil, ca 1210; ms. wall., ca 1210; MS2 (Ch. E. Stebbins, Romania 92, 1971, p. 1-36); DEAF : AlexisAloS.
- alia** : ANON. (1185*). *Roman d'Alexandre*. Ms. A; Dees : Vendée, Deux-Sèvres; compo. pic. et occ. orient., ca 1185; ms. poit., ca 1225; MS (ANON. 1937); DEAF : AlexArsL.

- amad** : ANON. (1210*). *Amadas et Ydoine*. Ms. P; Dees : Aisne; compo. a.n., ca 1210; ms. pic., ca 1289; CR3 (ANON. 1974); DEAF : AmYdR.
- amile** : ANON. (1200*). *Ami et Amile*. Dees : Haute-Marne; compo. nil, ca 1200; ms. lorr., ca 1275; MS1 (DEMBOWSKI 1969); DEAF : AmAmD.
- amo** : JACQUES D'AMIENS (1250*). *L'art d'amors et li remedes d'Amors*. Dees : Nord; compo. flandr., ca 1250; ms. pic. (?), ca 1300; MS (JACQUES D'AMIENS 1868); DEAF : BestAmOctT.
- amou** : ANON. (1250*). *Le bestiaire d'amour rimé*. Dees : Oise; compo. franc., ca 1250; ms. Nord, ca 1300; CR (THORDSTEIN 1941); DEAF : BestAmOctT.
- anth** : WAUCHIER DE DENAIN (1208*). *La vie de saint Antoine*. Dees : Haute-Marne; compo. wall., ca 1208; ms. pic., ca 1275; MS1 (transcr. inédite Th. Venckeleer).
- anti** : HUON DE MERY (1236*). *Le torneiment Anticrist*. Dees : [Haute-Marne]; compo. norm., ca 1236; ms. a.n., ca 1288; CR (HUON DE MERY 1976); DEAF : TournAntB.
- arr*** : NIL (nil). *La vie de saint Hylaire*. Dees : [Somme, Pas-de-Calais]; compo. nil, nil; ms. Arras, ca 1275; MS1 (transcr. inédite M. Raeymaeker); DEAF : SHiLR_A*.
- artch** : JEAN DE MEUN (1288*). *L'art de chevalerie*. Dees : Marne; compo. frc., ca 1288; ms. pic., 1340; CR2 (JEAN DE MEUN 1897); DEAF : JMeunVégR.
- athi** : ANON. (1200*). *Li romanz d'Athis et Prophilius*. Dees : Aisne; compo. pic., ca 1200; ms. frc., ca 1350; CR2 (HILKA 1912-1916); DEAF : AthisH.
- atre*** : ANON. (1250*). [PVR] *L'atre perilleux, v.1-820 B. Woledge, CFMA, Paris 1936 (N1) BN ffr 2168 fin 13e siècle*. Dees : [Somme, Pas-de-Calais]; compo. ouest, ca 1250; ms. pic., ca 1290; CR2 (B. Woledge, CFMA, Paris 1936); DEAF : AtreW.
- auc** : ANON. (1225*). *Aucassin et Nicolette*. Dees : Pas-de-Calais; compo. pic., ca 1225; ms. pic., ca 1290; MS1 (ROQUES 1955); DEAF : AucR³; ne contient que les parties en prose.
- aucchants** : ANON. (1225*). *Aucassin et Nicolette*. Dees : Pas-de-Calais; compo. pic., ca 1225; ms. pic., ca 1290; MS1 (ROQUES 1955); DEAF : AucR³; ne contient que les laisses chantées.
- avu*** : CORTEBARBE (1250*). *Les trois aveugles de Compiègne de Cortebarbe*. Dees : [Normandie]; compo. Oise, ca 1250; ms. frc., ca 1288; CR2 (G. Gougenheim, CFMA, Paris 1932); DEAF : TroisAvG.
- aye** : ANON. (1200*). *Orson de Beauvais*. Dees : [Normandie]; compo. pic./norm., ca 1200; ms. nil, ca 1300; CR1 (ANON. 1967); DEAF : AyeB.
- baisieux** : JACQUES DE BAISIEUX (1300*). [titre manquant]. Dees : Wallonie; compo. wall., ca 1300; ms. wall., ca 1300; MS (JACQUES DE BAISIEUX 1973); DEAF : JacBaisT.
- bar** : ANON. (1210*). *Le chevalier au barisel*. Dees : Somme, Pas-de-Calais; compo. pic., ca 1210; ms. pic., 1310; CR2 (ANON. 1955); DEAF : ChevBarAnL.
- barlaam** : ANON. (1217*). *L'histoire de Barlaam et Josaphat*. Dees : Haute-Marne; compo. champ., ca 1217; ms. Est, ca 1290; CR3 (MILLS 1973); DEAF : BalJosPr^{1M}.
- barril** : JOUHAM DE LA CHAPELE DE BLOIS (1218*). *Le conte dou barril*. Dees : Nièvre, Allier; compo. orl., ca 1218; ms. orl., ca 1317; MS (JOUHAM DE LA CHAPELE DE BLOIS 1932); DEAF : ChevBarBloisB.
- beati** : ANON. (1200*). *Dialogus anime conquerentis et rationis consolantis*. Dees : Vosges; compo. lorr., ca 1200; ms. lorr., ca 1200; MS2 (BONNARDOT 1876); DEAF : DialAmeB.
- beauv*** : PHILIPPE DE BEAUMANOIR (1283). *Les coutumes du Beauvoisis par Ph. de Beaumanoir, t.I, p. 356-441*. Dees : [Oise]; compo. Oise, 1283; ms. frc., ca 1250; MS3 (Comte Beugnot, Paris 1842); DEAF : BeaumCoutB_HarL.
- bel** : RENAUT DE BEAUJEU (1200*). *Le bel inconnu*. Dees : [Somme, Pas-de-Calais]; compo. Est, ca 1200; ms. hain., ca 1283; MS2 (RENAUT DE BEAUJEU 1929); DEAF : RenBeaujBelW².
- benoit** : BENOIT DE SAINTE-MAURE (1174*). *Chronique des ducs de Normandie*. Ms. T; Dees : Vendée, Deux-Sèvres; compo. poit., ca 1174; ms. poit., ca 1190; MS1 (BENOIT DE SAINTE-MAURE 1951-1967).
- bern2** : BERNARD DE CLAIRVAUX (1190*). *Sermons*. Dees : Moselle, Meurthe-et-Moselle; compo. lorr., ca 1190; ms. lorr., ca 1190; MS (BERNARD DE CLAIRVAUX 1894); DEAF : SBernAn2S.
- besant** : GUILLAUME LE CLERC DE NORMANDIE (1227). *Le besant de Dieu*. Compo. norm., 1227; ms. a.n., ca 1290; CR (GUILLAUME LE CLERC DE NORMANDIE 1973), DEAF : BesantR.
- best** : GUILLAUME LE CLERC DE NORMANDIE (1211*). *Bestiaire de le divine escripture*. Dees : Angleterre; compo. norm., ca 1211; ms. a.n., ca 1250; cr3 (GUILLAUME LE CLERC DE NORMANDIE 1892); DEAF : BestGuillR.
- bodo*** : ANON. (1200*). *La vie de saint Jean Bouche d'or*. Dees : [Somme, Pas-de-Calais]; compo. pic., ca 1200; ms. nil, nil; MS2 (A. Weber, Romania 6, 1877, p. 328-340); DEAF : SJeanBoucheW.

- bourg*** : ANON. (1310*). *Des poignes d'enfer, les deux chevaliers, enseignement moral*. Dees : [Franche-Comte]; compo. nil, ca 1310; ms. bourg., ca 1310; MS1 (P. Meyer, Romania 6, 1877, p. 1-46); DEAF : PoignEnfM*, DeuxChevM*, EnsMorM*.
- calen** : CALENDRE (1215*). *Des empereors de Rome*. Dees : Haute-Marne; compo. champ., 1215; ms. champ., ca 1235; MS2 (SETTEGAST 1879); DEAF : CalendreS.
- calex** : GUI (?) (1225). *La vie de sainte Catherine d'Alexandrie*. Ms. La Clayette; Dees : Région Parisienne; compo. pic., 1225; ms. nil, ca 1290; MS (TODD 1900); DEAF : SCathGuiT.
- cambrai** : ANON. (1190*). *Raoul de Cambrai*. Dees : Ardennes; compo. Nord-Est, ca 1190; ms. pic., ca 1225; CR1 (MEYER et LONGNON 1882); DEAF : RCambrM.
- carem** : ANON. (1217*). *Sermons de carême*. Dees : Wallonie; compo. liég., ca 1217; ms. liég., ca 1225; MS2 (PASQUET 1888); DEAF : SermCarP.
- carp*** : WAUCHIER DE DENAIN (1208*). *La vie de saint Hylaire*. Dees : [Nievre, Allier]; compo. wall., ca 1208; ms. pic., ca 1275; MS (transcr. inédite M. Raeymaekers); DEAF : SHiLR_C*.
- cass** : ANON. (1275*). *Le Romand de Cassidorus*. Dees : Marne; compo. pic., ca 1275; ms. frc., ca 1290; CR2 (PALERMO 1963); DEAF : CassidP.
- chaitH*** : MARIE DE FRANCE (1165*). *nil*. Dees : [Angleterre]; compo. agn., ca 1165; ms. agn., ca 1275; MS (K. Warnke, Halle 1900); DEAF : MarieChaitW2.
- chastoi*** : PIERRE ALPHONSE (1210*). [PVR] *Le chastoiment d'un pere a son fils, v.1- 3012 E.D. Montgomery Jr, Chapel Hill 1971 Brit.Mus. MS Additinal 10289 13e s*. Dees : [Normandie]; compo. norm., ca 1210; ms. norm., ca 1275; CR3 (E. D. Montgomery jr., Chapel Hill 1971); DEAF : ChastPereAM.
- chauvency** : JACQUES BRETTEL (1285*). *Le tournoi de Chauvency*. Ms. M; Dees : Haute-Marne; compo. Meuse, ca 1285; ms. loir., ca 1310; CR1 (JACQUES BRETTEL 1932); DEAF : BretTournD.
- chevreH*** : MARIE DE FRANCE (1165*). *nil*. Dees : [Angleterre]; compo. agn., ca 1165; ms. agn., ca 1275; MS (H. Suchier, Halle 1900); DEAF : MarieChievreS_H.
- chevres*** : MARIE DE FRANCE (1165*). *nil*. Dees : nil; compo. agn., ca 1165; ms. frc., ca 1300; MS (H. Suchier, Halle 1900); DEAF : MarieChievreS_S.
- chret1** : CHRÉTIEN DE TROYES (1180*). *Le Chevalier au lion & Perceval*. Dees : Hainaut; compo. champ. mérid., ca 1180; ms. champ., ca 1238; MS (transcr. inédite R. L. H. Lops); contenu présent ailleurs dans le corpus : incipit du *Chevalier au lion* (v. 1-998) dans yvf, et *Perceval* (v. 69-2000) dans perr.
- chret2** : CHRÉTIEN DE TROYES (1180*). *Le Chevalier au lion & Perceval*. Ms. H (*Le Chevalier au lion*) & Ms. A (*Perceval*); Dees : Haute-Marne; compo. champ. mérid., ca 1180; ms. champ., ca 1235; MS (transcr. inédite R. L. H. Lops & CHRÉTIEN DE TROYES 1975); contenu présent ailleurs dans le corpus : incipit du *Chevalier au lion* (v. 1-998) dans yvh, et *Perceval* (v. 1-2000) dans pera; DEAF (*Perceval*) : PercL.
- chro** : BENOIT DE SAINTE-MAURE (1174*). *Chronique des ducs de Normandie*. Ms. [Tours, Bibl. Munic. 903]; Dees : [Vendée, Deux-Sèvres]; compo. poit., ca 1174; ms. a.n., ca 1217; CR2 (BENOIT DE SAINTE-MAURE 1951-1967); DEAF : BenDucF.
- clari2** : ROBERT DE CLARI (1217*). *La conquête de Constantinople*. Dees : Somme, Pas-de-Calais; compo. pic., ca 1217; ms. pic., 1300; MS (ROBERT DE CLARI 1956); DEAF : RobClaril.
- clef*** : ANON. (1280). *La clef d'amors*. Dees : [Normandie]; compo. norm., 1280; ms. pic., ca 1317; CR (A. Doutrepont, Normannica V, Halle 1890); DEAF : ClefD.
- cleom*** : ADENET LE ROI (1285). *Les oeuvres d'Adenet le Roi, t.V Cleomades, ms. A, v. 1-2566*. Dees : [Oise]; compo. flandr., 1285; ms. frc., ca 1290; CR1 (A. Henry, Bruxelles 1971); DEAF : CleomH.
- coinci*** : GAUTIER DE COINCY (1226*). *Les miracles de Nostre Dame de Gautier de Coinci, ms. L, p. 5-176*. Dees : [Marne]; compo. nord, ca 1226; ms. pic. merid., ca 1310; MS1 (F. Koenig, TLF, Genève 1955); DEAF : CoincyI*.
- compoit** : ANON. (ca 1225). *Roman du Comte de Poitiers*. Compo. pic., ca 1225; ms. pic., ca 1310; MS (MALMBERG 1940); DEAF : ComtePoitM.
- conperc*** : GERBERT DE MONTREUIL (1238*). *La continuation de Perceval, ms. A, v. 1-1520*. Dees : [Somme, Pas-de-Calais]; compo. pic., ca 1238; ms. pic., ca 1210; MS1 (M. Williams, CFMA, Paris 1922); DEAF : ContPerc_4_TW*.
- contre** : ANON. (1322*). *Le Roman de Renart le contrefait, fin de la 6^e et dernière branche*. Dees : Aube; compo. champ. mérid., ca 1322; ms. nil, ca 1337; CR2 (LEMAÎTRE et RAYNAUD 1914); DEAF : RenContr¹R.

- contro** : ANON. (1200*). *La seconde continuation de Perceval*. Dees : Aube; compo. nil, ca 1200; ms. champ. mérid., ca 1225; CR1 (ROACH 1971); DEAF : ContPerc²R.
- cordres** : ANON. (1200*). *La prise de Cordres et de Seville*. Dees : Meuse; compo. lorr., ca 1200; ms. Meuse, ca 1262; MS2 (DENSUSIANU 1965 [1896]); DEAF : PriseCordD.
- cou** : ANON. (1210*). *Courtois d'Arras*. Dees : Hainaut; compo. art., ca 1210; ms. pic., ca 1285; cr1 (FARAL 1922).
- darm** : ANON. (1275*). *Médecinaire Liégeois du XIII^e siècle*. Dees : Wallonie; compo. liég., ca 1275; ms. wall, ca 1275; CR1 (HAUST 1941); DEAF : MédLiégH.
- desp** : ANON. (1238*). *Li chevaliers as deus espees*. Dees : Pas-de-Calais; compo. pic., ca 1238; ms. Pas-de-Calais, ca 1300; CR2 (FOERSTER 1877); DEAF : CheviiEspF.
- deusamH** : MARIE DE FRANCE (1165*). *Les Deus Amanz*. Ms. H; Dees : nil; compo. a.n., ca 1165; ms. a.n., ca 1275; MS (transcr. inédite sans transcripateur indiqué).
- deusamS** : MARIE DE FRANCE (1165*). *Les Deus Amanz*. Ms. S; Dees : nil; compo. a.n., ca 1165; ms. frc., ca 1300; MS (transcr. inédite sans transcripateur indiqué).
- dole*** : ANON. (1209*). *Le roman de la rose ou de Guillaume de Dole, v. 1-803*. Dees : [Haute-Marne]; compo. pic., ca 1209; ms. frc., ca 1290; CR3 (G. Servois, SATF, Paris 1893); DEAF : GuillDoleS.
- durm*** : ANON. (1217*). [PVR] *Durmart le Galois, Tome I, v.1-800* J. Gildea, Villanova 1965 Ms Berne 113 nv. Dees : [Wallonie]; compo. pic., ca 1217; ms. bourg., ca 1290; CR1 (J. Gildea, Villanova 1965); DEAF : DurmG.
- edmond** : MATTHIEU PARIS (1255*). *La vie de saint Edmond*. Dees : nil; compo. a.n., ca 1255; ms. a.n., ca 1280; MS2 (J.-A. BAKER 1929); DEAF : SEdmCantB.
- edmund** : DENIS PIRAMUS (1195*). *La Vie Seint Edmund le Rei*. Dees : [Angleterre]; compo. a.n., ca 1195; ms. a.n., ca 1325; MS2 (DENIS PIRAMUS 1935); DEAF : EdmK.
- egip*** : RUTEBEUF (1263*). *Sainte Marie l'Egyptienne, v. 1-813*. Dees : [Aisne]; compo. champ. merid., ca 1263; ms. frc., ca 1288; CR2 (E. Faral et J. Bastin, Paris 1959); DEAF : RutebF.
- elid** : MARIE DE FRANCE (ca 1165). *Lais*. Compo. NO, ca 1165; ms. a.n., ca 1275; MS (MARIE DE FRANCE 1947).
- elie*** : ANON. (1190*). *Elie de Saint Gille*. Dees : [Somme, Pas-de-Calais]; compo. pic., ca 1190; ms. pic., ca 1275; nil (transcr. inédite P. Bloem); DEAF : ELieB*.
- eneas** : ANON. (1160*). *Éneas*. Ms. A; Dees : Haute-Marne; compo. norm., ca 1160; ms. lorr., ca 1200; CR2 (SALVERDA DE GRAVE 1964-68).
- enf*** : ANON. (1275*). *The Old French Evangile de l'Enfance, ms. G*. Dees : [Normandie]; compo. Ouest, ca 1275; ms. pic., ca 1317; MS (M.B. McCann Boulton, Ann Arbor 1980); DEAF : EvEnfB.
- epee** : ANON. (1200*). *Le chevalier à l'épée*. Dees : Nièvre, Allier; compo. bourg., ca 1200; ms. bourg. sept., ca 1310; MS1 (JOHNSTON et OWEN 1972); DEAF : ChevEspJ.
- equiH*** : MARIE DE FRANCE (1165*). *nil*. Dees : [Angleterre]; compo. agn., ca 1165; ms. agn., ca 1275; MS (H. Suchier, Halle 1900 :); DEAF : MarieEquitS_H.
- equiS*** : MARIE DE FRANCE (1165*). *nil*. Dees : nil; compo. agn., ca 1165; ms. frc., ca 1300; MS (H. Suchier, Halle 1900 :); DEAF : MarieEquitS_S.
- eustache*** : BENOÎT (1210*). *La vie de Saint Eustache, ms. M*. Dees : [Normandie]; compo. norm., ca 1210; ms. norm., ca 1250; MS2 (H. Petersen, CFMA, Paris 1928); DEAF : SEust2P.
- evrat1** : EVRAT (1198*). *Genèse*. Dees : nil; compo. champ., ca 1198; ms. lorr., ca 1210; MS1-2 (transcr. inédite W. Boers), suite (?) dans evratC2.
- evratC2** : EVRAT (1198). *Genèse*. Dees : nil; compo. champ., 1198; ms. lorr., ca 1210; MS1 (transcr. inédite W. Boers), suite (?) de evrat1.
- fab4c** : ANON. (1200*). *Fabliau n°4*. Ms. C; Dees : Normandie; compo. pic., ca 1200; ms. norm., ca 1300; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).
- fab4e** : ANON. (1200*). *Fabliau n°4*. Ms. E; Dees : Nièvre, Allier; compo. pic., ca 1200; ms. frc., ca 1290; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).
- fab4f** : ANON. (1200*). *Fabliau n°4*. Ms. F; Dees : Somme, Pas-de-Calais; compo. pic., ca 1200; ms. nil, ca 1300; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).
- fab4a** : ANON. (1250*). *Fabliaux n°1, 2, 4, 23, 29*. Ms. A; Dees : Normandie; compo. pic., ca 1250; ms. frc., ca 1288; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).
- fab4b** : ANON. (1250*). *Fabliaux n°2, 4*. Ms. B; Dees : Haute-Marne; compo. pic., ca 1250; ms. bourg. sept., ca 1310; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).

- fabd** : ANON. (1250*). *Fabliaux n°2, 4*. Ms. D; Dees : Haute-Marne; compo. pic., ca 1250; ms. frc., ca 1290; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).
- fabj** : ANON. (1250*). *Fabliaux n°2, 4*. Ms. J; Dees : Somme, Pas-de-Calais; compo. pic., ca 1250; ms. pic., ca 1285; MS (NOOMEN et VAN DEN BOOGAARD 1983-1998).
- falesA** : MARIE DE FRANCE (1180*). *Fables*. Ms. A; Dees : [Angleterre]; compo. NO, ca 1180; ms. a.n., ca 1250; MS (transcription inédite).
- falesB** : MARIE DE FRANCE (1180*). *Fables*. Ms. B; Dees : [Angleterre]; compo. NO, ca 1180; ms. a.n., ca 1290; MS (transcription inédite).
- falesC** : MARIE DE FRANCE (1180*). *Fables*. Ms. C; Dees : nil; compo. NO, ca 1180; ms. lorr., ca 1275; MS (transcription inédite).
- falesE1** : MARIE DE FRANCE (1180*). *Fables*. Ms. E; Dees : [Angleterre]; compo. NO, ca 1180; ms. a.n., ca 1225; MS (transcription inédite).
- falesK** : MARIE DE FRANCE (1180*). *Fables*. Ms. K; Dees : nil; compo. NO, ca 1180; ms. nil, ca 1310; MS (transcription inédite).
- falesL** : MARIE DE FRANCE (1180*). *Fables*. Ms. L; Dees : nil; compo. NO, ca 1180; ms. nil, ca 1290; MS (transcription inédite).
- falesM** : MARIE DE FRANCE (1180*). *Fables*. Ms. M; Dees : nil; compo. NO, ca 1180; ms. wall., ca 1300; MS (transcription inédite).
- falesT** : MARIE DE FRANCE (1180*). *Fables*. Ms. T; Dees [Angleterre]; compo. NO, ca 1180; ms. champ., ca 1287; MS (transcription inédite).
- falesY** : MARIE DE FRANCE (1180*). *Fables*. Ms. Y; Dees [Angleterre]; compo. NO, ca 1180; ms. norm. et a.n., ca 1225; MS (transcription inédite).
- falesZ** : MARIE DE FRANCE (1180*). *Fables*. Ms. Z; Dees nil; compo. NO, ca 1180; ms. nil, ca 1400; MS (transcription inédite).
- fauc*** : ANON. (1250*). *Li dis du faucon*. Dees : [Somme, Pas-de-Calais]; compo. pic., ca 1250; ms. pic., 1295; CR (transcr. inédite W. van Hoecke); DEAF : ComparFaucVH*.
- fetrom** : ANON. (1213*). *Li fet des Romains*. Ms. V; Dees : Yonne; compo. frc., ca 1213; ms. bourg., ca 1290; CR1 (FLUTRE et SNEIJERS DE VOGEL 1935); DEAF : FetRomF¹.
- feu*** : ADAM LE BOSSU (DE LA HALLE) (1276). *Le jeu de la feuillée d'Adam le Bossu*. Dees : [Somme, Pas-de-Calais]; compo. art., 1276; ms. pic., 1295; CR3 (E. Langlois, CFMA, Paris 1951); DEAF : AdHaleFeuilLL.
- fierens** : ANON. (1489). *Cartulaire de l'hôpital Saint-Jacques de Tournai*. Dees : nil; compo. Tournai, ca 1489; ms. Tournai, ca 1489; MS (transcr. inédite non identifiée).
- flo** : ANON. (1213*). *Florence de Rome*. Dees : Nièvre, Allier; compo. pic., ca 1213; ms. Est, ca 1300; CR1 (WALLENSKÖLD 1907-1909); DEAF : FlorenceW.
- floov** : ANON. (1190*). *Floovant, Chanson de Geste du xiii^e siècle*. Dees : Bourgogne; compo. Sud-Est, ca 1190; ms. bourg., ca 1325; MS1 (ANDOLF 1965); DEAF : FloovA.
- gar** : ANON. (1275*). *Le Garçon et l'aveugle*. Dees : Aisne; compo. pic., ca 1275; ms. pic., ca 1275; CR2 (ANON. 1969 [1921]); DEAF : GarçAvR².
- gepa** : GEOFFROY DE PARIS (1316). *Chronique métrique*. Dees : Normandie; compo. norm., 1316; ms. nil, ca 1318; MS (GEOFFROY DE PARIS 1956); DEAF : GeoffrParChronD.
- gerv*** : GERVAISE (1210*). *Le bestiaire de Gervaise*. Dees : [Bourgogne]; compo. bourg., ca 1210; ms. bourg., ca 1275; MS1 (P.Meyer, Romania 1, 1872, p. 420-443); DEAF : BestGervM.
- gibv*** : GUILLAUME DE BERNEVILLE (1183*). *La vie de Saint Gilles par Guillaume de Berneville*. Dees : [Angleterre]; compo. agn., ca 1183; ms. agn., ca 1225; CR1 (G. Paris et A. Bos, SATF, Paris 1881 (reprint 1966)); DEAF : SGillesP.
- gir*** : BERTRAND DE BAR-SUR-AUBE (1210*). *Girart de Vienne, par Bertrand de Bar-sur-Aube, v. 1-1024*. Dees : [Nievre, Allier]; compo. champ. merid., ca 1210; ms. bourg., ca 1270; CR2 (W. van Emden, SATF, Paris 1977); DEAF : GirVianeE.
- gorm** : ANON. (1125*). *Gormont et Isembart*. Dees : Angleterre; compo. frc., ca 1125; ms. a.n., ca 1213; MS2 (BAYOT 1931); DEAF : GormB.
- graal*** : ROBERT DE BORON (1196*). *Le roman de l'estoire dou graal de Robert de Boron, v. 1-850*. Dees : [Aisne]; compo. frcomt., ca 1196; ms. pic., ca 1290; MS1 (W.A. Nitze, CFMA, Paris 1927); DEAF : SGraaLIJJoN.

- greg2** : JEAN LE DIACRE (1215). *La vie de Saint Gregoire le Grand*. Dees : [Angleterre]; compo. a.n., 1215; ms. a.n., ca 1215; CR (JEAN LE DIACRE 1883); DEAF : AngVieGregM.
- guib*** : ANON. (1210*). *Guibert d'Andrenas, v. 1-849*. Dees : [Nievre, Allier]; compo. frc., ca 1210; ms. bourg., ca 1270; CR3 (J. Melander, Paris 1922); DEAF : GuibAndrM.
- guigH** : MARIE DE FRANCE (1165*). *Le Lai de Guigemar*. Ms. H; Dees : nil; compo. N.-O., ca 1165; ms. a.n., ca 1275; CR (transcr. inédite sans transcripateur indiqué).
- guigP** : MARIE DE FRANCE (1165*). *Le Lai de Guigemar*. Ms. P; Dees : nil; compo. N.-O., ca 1165; ms. pic., ca 1290; nil (transcr. inédite sans transcripateur indiqué).
- guigS** : MARIE DE FRANCE (1165*). *Le Lai de Guigemar*. Ms. S; Dees : nil; compo. N.-O., ca 1165; ms. frc., ca 1300; nil (transcr. inédite sans transcripateur indiqué).
- guil*** : ANON. (1223*). [PVR] *L'Histoire de Guillaume le Marechal, Comte de Striguil et de Pembroke v.1-804 P. Meyer, Paris 1891-94 t.I bibl. Phillips Cheltenham n verm*. Dees : [Vendee, Deux-Sevres]; compo. norm., ca 1223; ms. agn., ca 1237; CR2 (P. Meyer, Paris 1891-94); DEAF : GuilMarM.
- hard*** : SAINT BERNARD DE CLAIRVAUX (1190*). *Li Sermon Saint Bernart*. Dees : [Moselle, Meurthe-et-Moselle]; compo. lorr., ca 1190; ms. Metz, ca 1190; CR2 (W. Foerster, Rom.Forsch. II, 1886, p. 1-15); DEAF : SBernAn1F.
- helc** : ANON. (1285*). *Le roman de Helcanus*. Dees : région parisienne; compo. pic., ca 1285; ms. frc., ca 1290; CR2 (NIEDZIELSKI 1966); DEAF : HelcanusN.
- herm*** : HERMAN DE VALENCIENNES (1190*). *Li romanz de Dieu et de sa mere d'Herman de Valenciennes, v. 1-2224*. Dees : [Haute-Marne]; compo. pic., ca 1190; ms. lorr., ca 1290; CR (I. Spiele, Leiden 1975); DEAF : HermValS.
- hista** : WAUCHIER DE DENAIN (1213*). *Histoire ancienne jusqu'à César (Estoires Rogier)*. Ms. [BFN fr. 20125 & Vienne Nat. Bibl. 2576]; Dees : [Haute-Marne]; compo. Belgique, ca 1213; ms. Nord-Est, ca 1287; MS1 (WAUCHIER DE DENAIN 1995); DEAF : HistAncV.
- hue*** : HUE ARCHEVESQUE (1250*). *Les dits de Hue Archevesque*. Dees : [Normandie]; compo. norm., ca 1250; ms. frc., ca 1287; MS (A. Héron, Rouen 1885); DEAF : HueArchH*.
- hyla** : WAUCHIER DE DENAIN (1208*). *La vie de saint Hylaire*. Dees : Somme, Pas-de-Calais; compo. Belgique, ca 1208; ms. pic., ca 1275; MS1 (transcr. inédite Th. Venckeleer).
- ipo** : HUE DE ROTELANDE (ca 1185). *Ipomedon*. Compo. a.n., ca 1185; ms. a.n., ca 1190; CR (transcription inédite).
- jaco** : PIERRE DE BEAUVAIS (1212). *Translacion mon seigneur saint Jaque*. Dees : Aube; compo. Beauvais, 1212; ms. champ., ca 1290; CR (BERKEY 1965); DEAF : PBeauvJacB.
- jean*** : HENRI DE VALENCIENNES (1225*). [PVR] *La vie de saint Jean l'Evangeliste E. Westberg, Uppsala 1943 (M) Madrid BN 9446 mil. 13e s.* Dees : [Aube]; compo. hain., ca 1225; ms. norm., ca 1250; CR (E. Westberg, Uppsala 1943); DEAF : SJeanEvW.
- joinv*** : JEAN DE JOINVILLE (1307*). *La vie de saint Louis de J. de Joinville, par. 1-165*. Dees : [Region parisienne]; compo. champ., ca 1307; ms. Nord-Est, ca 1335; CR1 (N.L. Corbett, Sherbrooke 1977); DEAF : JoinvC.
- jongl*** : D'APRÈS GEUFROI DE PARIS (1243). *La Passion des Jongleurs, v. 1-1501*. Dees : [Region parisienne]; compo. frc., 1243; ms. frc., ca 1275; MS A.J.A. Perry, Paris 1981); DEAF : PassJonglGP.
- jouf** : ANON. (1250*). *Joufroi de Poitiers*. Dees : Haute-Marne; compo. SE, ca 1250; ms. Est., ca 1300; MS3 (FAY et GRIGSBY 1972); DEAF : JoufrF.
- juda*** : ANON. (1285). *La chevalerie de Judas Macabé, ms. E, v. 1-1546*. Dees : [Somme, Pas-de-Calais]; compo. pic., 1285; ms. Nord, ca 1290; MS1 (J.R. Smeets, Assen 1955); DEAF : MaccabES.
- juise2** : ANON. (1137*). *Li ver del Juise*. Dees : Wallonie; compo. lié., ca 1137; ms. wall., ca 1210; CR2 (RANKKA 1982); DEAF : JuiseR.
- kathe*** : ANON. (1250*). *La passion saynte Katherine*. Dees : [Franche-Comte]; compo. poit., ca 1250; ms. traits frpr., ca 1325; MS1 (E.C. Fawtier-Jones, Romania LXIII, 1932, p.211-217); DEAF : SCathJonglF.
- lac** : ANON. (1237*). *Lancelot do Lac*. Ms. Ao; Dees : Haute-Marne; compo. nil, ca 1220; ms. Est, ca 1237; CR1 (KENNEDY 1980); DEAF : LancPrK.
- lanc*** : ANON. (1220*). *Lancelot, roman en prose du XIIIe siècle ms. A, par. 1.1-4.2*. Dees : [Somme, Pas-de-Calais]; compo. nil, ca 1220; ms. pic., ca 1275; CR (A. Micha, Genève 1978); DEAF : LancPrMLA.
- lancpr** : ANON. (1220*). *Lancelot*. Ms. S; Dees : [Somme, Pas-de-Calais]; compo. nil, ca 1220; ms. pic., ca 1316; qualité nil (MICHA 1978-1983); DEAF : LancPrM.
- lanvalc** : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. C; Dees : Angleterre; compo. N.-O., ca 1165; ms. a.n., ca 1290; MS2 (MARIE DE FRANCE 1958); DEAF : MarieLanvR².

- lanvalH** : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. H; Dees : nil; compo. N.-O., ca 1165; ms. a.n., ca 1275; MS2 (MARIE DE FRANCE 1958); DEAF : MarieLanvR².
- lanvalP** : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. P; Dees : nil; compo. N.-O., ca 1165; ms. pic., ca 1290; MS2 (MARIE DE FRANCE 1958); DEAF : MarieLanvR².
- lanvalS** : MARIE DE FRANCE (1165*). *Le lai de Lanval*. Ms. S; Dees : nil; compo. N.-O., ca 1165; ms. frc., ca 1300; MS2 (MARIE DE FRANCE 1958); DEAF : MarieLanvR².
- laustH*** : MARIE DE FRANCE (1165*). *nil*. Dees : [Angleterre]; compo. Nord-Ouest, ca 1165; ms. agn., ca 1275; nil (Warnke, Halle 1925); DEAF : MarieLaustW3*.
- lechS*** : ANON. (1210*). *Le Lai du Lecheor*. Dees : nil; compo. frc., ca 1210; ms. frc., ca 1300; CR1 (?); DEAF : nil.
- lin** : ROBERT DE BORON (1213*). *Merlin*. Ms. A; Dees : Yonne; compo. Sud-Est, ca 1213; ms. bourg., ca 1250; CR2 (MICHU 1980); DEAF : MerlinM.
- livre** : ESTIENNE DE FOGIERES (ca 1176). *Livre des manières*. Compo. Ouest, ca 1176; ms. Ouest, ca 1217; MS1 (ESTIENNE DE FOGIERES 1979); DEAF : EstFougL.
- loth*** : ANON. (1365). *Lothringischer Psalter, p. 1-23*. Dees : [Meuse]; compo. lorr., 1365; ms. lorr., ca 1375; MS (F. Apfelstedt, Wiesbaden 1881); DEAF : PsLorra.
- loys*** : GUILLAUME DE SAINT-PATHUS (1298*). *Les miracles de saint Louis de Guillaume de St Pathus, par. 1-54*. Dees : [Normandie]; compo. multiple, ca 1298; ms. nil, ca 1310; MS1 (P.B. Fay, CFMA, Paris 1932); DEAF : SLouisPathMirF.
- mace** : MACÉ DE LA CHARITÉ (1300*). *La Bible*. Dees : Nièvre, Allier; compo. Centre, ca 1300; ms. Berry, 1343; CR (MACÉ DE LA CHARITÉ 1977); DEAF : BibleMacéV.
- malk** : JEHAN MALKARAUME (1283*). *La bible de Jehan Malkaraume*. Dees : Marne; compo. lorr., ca 1283; ms. lorr., id.; CR2 (JEHAN MALKARAUME 1987); DEAF : BibleMalkS.
- marga** : WACE (1135*). *La vie de Sainte Marguerite de Wace*. Ms. M; Dees : Haute-Marne; compo. norm., ca 1135; ms. art., 1267; CR3 (WACE 1932a); DEAF : WaceMargaF.
- martin1** : PÉAN GÂTINEAU (1125*). *Vie de saint Martin*. Dees : Nièvre, Allier; compo. tourr., ca 1125; ms. bourg., ca 1290; MS (PÉAN GÂTINEAU 1896); DEAF : PeanGatS1.
- martin2** : PÉAN GÂTINEAU (1125*). *Vie de saint Martin*. Dees : Vendée, Deux-Sèvres; compo. tourr., ca 1125; ms. poit., ca 1290; MS (PÉAN GÂTINEAU 1896); DEAF : PeanGatS1.
- martin3** : PÉAN GÂTINEAU (1125*). *Vie de saint Martin*. Dees : Indre, Cher; compo. tourr., ca 1125; ms. berr., ca 1290; MS (PÉAN GÂTINEAU 1896); DEAF : PeanGatS1.
- maur** : BENOÎT DE SAINTE-MAURE (1170*). *Le roman de Troie*. Dees : [Vendee, Deux-Sevres]; compo. poit., ca 1170; ms. a.n., ca 1210; CR3 (BENOÎT DE SAINTE-MAURE 1904); DEAF (ToDo) : BenTroieC.
- mede*** : ANON. (1275*). *Remèdes populaires du moyen âge*. Dees : [Somme, Pas-de-Calais]; compo. Nord, ca 1275; ms. pic., ca 1290; MS1 (A. Salmon, Paris 1891, p.253-266); DEAF : RecMédCambraiS*.
- merlin*** : ROBERT DE BORON (1213*). *Merlin, roman en prose du XIIIe siècle, p. 1-29*. Dees : [Somme, Pas-de-Calais]; compo. Sud-Est, ca 1213; ms. pic., ca 1310; CR1 (G. Paris et J. Ulrich, SATF, Paris 1886, p.1-29); DEAF : MerlinP.
- merm*** : PIERRE DE BEUVAIS (1200*). *Le bestiaire de Pierre de Beauvais (version courte)*. Dees : [Marne]; compo. pic., ca 1200; ms. Marne, ca 1290; CR3 (G.R. Mermier, Paris 1977); DEAF : BestPierre1M.
- meun*** : JEAN DE MEUNG (1230*). *Le Roman de la Rose de Jean de Meun, v. 4029-8226*. Dees : [Region parisienne]; compo. traits orl., ca 1230; ms. orl., ca 1285; CR2 (F. Lecoy, CFMA, 3 vol., Paris 1965); DEAF : RoseLLec.
- michel*** : GUILLAUME DE SAINT-PAIR (1163*). [PVR] *Le roman du Mont Saint-Michel par Guillaume de Saint-Pair, v.1-796 Francisque Michel, Caen 1856 nv nv*. Dees : [Normandie]; compo. Nord-Ouest, ca 1163; ms. norm., ca 1275; MS (Francisque Michel, Caen 1856); DEAF : ChronSMiche1M.
- milons*** : MARIE DE FRANCE (1165*). *nil*. Dees : nil; compo. Nord-Ouest, ca 1165; ms. frc., ca 1300; CR1 (transcr. inédite sans transcripateur indiqué); DEAF : MarieMil_S.
- milunH*** : MARIE DE FRANCE (1165*). *nil*. Dees : [Angleterre]; compo. Nord-Ouest, ca 1165; ms. agn., ca 1275; MS (?); DEAF : MarieMil_H.
- mir** : GAUTIER DE COINCI (1224*). *Les miracles de Notre Dame de Soissons*. Ms. R; Dees : Aisne; compo. Soissonnais, ca 1224; ms. pic. mérid., ca 1310; CR (GAUTIER DE COINCI 1966); DEAF : CoincyI10K.
- miro** : ROBERT DE L'OMME (1266). *Le miroir de vie et de mort*. Ms. G; Dees : Somme, Pas-de-Calais; compo. pic., 1266; ms. art., 1277; MS1 (LÅNGFORS 1921, 1924); DEAF : Rob0mMirL.

- modw** : ANON. (1230*). *Saint Modwenna*. Ms. O; Dees : Angleterre; compo. a.n., ca 1230; ms. a.n., ca 1250; CR (A. T. BAKER et BELL 1947); DEAF : [ModvB²](#).
- moral** : ANON. (1200*). *Le poème moral, traité de la Vie Chrétienne*. V. 1-2320; Dees : Wallonie; compo. wall., ca 1200; ms. wall., ca 1210; CR1 (BAYOT 1929); DEAF : [PoèmeMorB](#).
- moral2** : ANON. (1200*). *Le poème moral, traité de la Vie Chrétienne*. V. 2321-fin; Dees : Wallonie; compo. wall., ca 1200; ms. wall., ca 1311; CR1 (BAYOT 1929); DEAF : [PoèmeMorB](#).
- mortartu*** : ANON. (1213*). *La mort le roi Artu, par. 1-35*. Dees : [Region parisienne]; compo. nil, ca 1213; ms. frc., ca 1290; CR (J. Frappier, TLF, Genève/Paris 1964); DEAF : [MortArtuF3*](#).
- mous*** : PHILIPPE MOUSKES (1243). *Chronique rimée de Philippe Mouskes, v. 12135-14650*. Dees : [Hainaut]; compo. hain., 1243; ms. hain., ca 1275; MS3 (Baron De Reiffenberg, Bruxelles 1838); DEAF : [MousketR](#).
- mrgr1** : WACE (1135*). *Vie de Sainte Marguerite*. Ms. M; Dees : [Haute-Marne]; compo. norm., ca 1135; ms. tour., ca 1238; CR3 (WACE 1932b); DEAF : [WaceMarmF](#).
- mule*** : PAIEN DE MAISIÈRES (1200*). *La mule sans frein, dans : Two old french Gauvain romances*. Dees : [Nievre, Allier]; compo. bourg., ca 1200; ms. bourg. sept., ca 1310; MS1 (R.C. Johnston and D.D.R. Owen, Edinburgh/London 1972); DEAF : [Mu1eJ](#).
- myst** : ANON. (1175*). *Le mystère d'Adam*. Dees : [Angleterre]; compo. traits occ. et a.n., ca 1175; ms. tour., ca 1238; MS2 (ANON. 1963); DEAF : [AdamA](#).
- narcA** : ANON. (1165*). *Narcisus*. Dees : nil; compo. norm., ca 1165; ms. frc., ca 1287; MS (PELAN et SPENCE 1964); DEAF (ToDo) : [NarcisusP](#).
- narcB** : ANON. (1165*). *Narcisus*. Dees : nil; compo. norm., ca 1165; ms. frc., ca 1290; MS (PELAN et SPENCE 1964); DEAF (ToDo) : [NarcisusP](#).
- narcC** : ANON. (1165*). *Narcisus*. Ms. C; Dees nil; compo. norm., ca 1165; ms. pic., ca 1290; CR1 (PELAN et SPENCE 1964); DEAF : [NarcisusP](#).
- narcD** : ANON. (1165*). *Narcisus*. Dees : nil; compo. norm., ca 1165; ms. norm., ca 1300; MS (PELAN et SPENCE 1964); DEAF (ToDo) : [NarcisusP](#).
- narcE*** : ANON. (1165*). *nil*. Dees : nil; compo. norm., ca 1165; ms. art., 1267; MS (M. Thiry-Stassin et M. Tyssens, Narcisse, Paris 1976); DEAF : [NarcisusT](#).
- ndchar** : JEAN LE MARCHANZ (1157*). *Miracles de Notre Dame de Chartres*. Dees : Orléanais; compo. Chartres, ca 1157; ms. Chartres, ca 1350; MS2 (JEAN LE MARCHANZ 1973); DEAF : [MirNDChartrK](#).
- nee1e*** : PERROT DE NESLE (1289). *Peros von Neele's gereimte Inhaltsangabe*. Dees : [Somme, Pas-de-Calais]; compo. pic., 1289; ms. pic., 1289; MS (L. Jordan, Romanische Forschungen XVI 1904, 735-756); DEAF : [PerNeslesTabJ](#).
- nic** : JEHAN BODEL (1197*). *Le jeu de Saint Nicolas*. Dees : Somme, Pas-de-Calais; compo. art., ca 1197; ms. Arras, ca 1295; MS1 (JEHAN BODEL 1965); DEAF : [BodelNicJ²](#).
- nicb*** : PSEUDO-NICOMÈDE (1225*). *L'Évangile de Nicodème, tradition B*. Dees : [Marne]; compo. nil, ca 1225; ms. flandr., ca 1288; MS1 (A.E. Ford, Genève 1973); DEAF : [EvNicPrBF](#).
- nicoa** : PSEUDO-NICODÈME (1225*). *L'Évangile de Nicodème*. Ms. A; Dees : [Angleterre]; compo. nil, ca 1225; ms. a.n., ca 1290; MS1 (PSEUDO-NICODÈME 1973); DEAF : [EvNicPraF](#).
- nima1** : ANON. (1150*). *Le charroi de Nîmes*. Ms. A1; Dees : Haute-Marne; compo. nil, ca 1150; ms. frc., ca 1263; MS1 (transcr. inédite L. Schoesler).
- nima2** : ANON. (1150*). *Le charroi de Nîmes*. Ms. A2; Dees : Nièvre, Allier; compo. nil, ca 1150; ms. frc., ca 1263; MS1 (transcr. inédite L. Schoesler).
- nima3** : ANON. (1150*). *Le charroi de Nîmes*. Ms. A3; Dees : Nièvre, Allier; compo. nil, ca 1150; ms. lorr., ca 1325; MS1 (transcr. inédite L. Schoesler).
- nima4** : ANON. (1150*). *Le charroi de Nîmes*. Ms. A4; Dees : Nièvre, Allier; compo. nil, ca 1150; ms. frc., ca 1283; MS1 (transcr. inédite L. Schoesler).
- nimafrag*** : ANON. (1150*). *Le charroi de Nîmes, ms. Afragment*. Dees : [Haute-Marne]; compo. nil, ca 1150; ms. nil, ca 1250; MS1 (transcr. inédite L. Schoesler); DEAF : [CharroiSch_fragm*](#).
- nimb1** : ANON. (1150*). *Le charroi de Nîmes*. Ms. B1; Dees : Aisne; compo. nil, ca 1150; ms. Paris, ca 1335; MS1 (transcr. inédite L. Schoesler).
- nimb2** : ANON. (1150*). *Le charroi de Nîmes*. Ms. B2; Dees : Normandie; compo. nil, ca 1150; ms. Paris, ca 1335; MS1 (transcr. inédite L. Schoesler).

- nimc** : ANON. (1150*). *Le charroi de Nîmes*. Ms. C; Dees : Somme, Pas-de-Calais; compo. nil, ca 1150; ms. art., 1295; MS1 (transcr. inédite L. Schoesler).
- nimd** : ANON. (1150*). *Le charroi de Nîmes*. Ms. D; Dees : Meuse; compo. nil, ca 1150; ms. lorr. mérid., ca 1275; MS1 (transcr. inédite L. Schoesler).
- nimfrag** : ANON. (1150*). *Le charroi de Nîmes*. Ms. 'Afragment'; Dees : Haute-Marne; compo. nil, ca 1150; ms. nil., 1250; MS1 (transcr. inédite L. Schoesler).
- nouvel*** : JACQUEMART GIELEE (1290*). *Renart le Nouvel par Jacquemart Gielee, v. 1-3026*. Dees : [Somme, Pas-de-Calais]; compo. Lille, ca 1290; ms. Arras, 1295; MS (R. Roussel, SATF, Paris 1961); DEAF : RenNouvR.
- oakbook** : ANON. (s. d.). [*manquant*]. [*manquant*] MS (SMIRKE 1859).
- oct*** : ANON. (1275*). *Octavian, v. 1-804*. Dees : [Oise]; compo. pic., ca 1275; ms. pic., ca 1290; MS2 (K. Vollmoeller, Wiesbaden 1883); DEAF : FlorOct0ctV.
- ombre** : JEAN RENART (1221*). *Le Lai de l'Ombre*. Dees : [Marne]; compo. pic., ca 1221; ms. frc, ca 1288; CR1 (JEAN RENART 1913); DEAF : OmbreB¹.
- or** : RENAULT (1200*). *La vie de saint Jehan Bouche d'Or*. Dees : Pas-de-Calais; compo. pic., ca 1200; ms. pic., 1370; CR1 (RENAULT 1931); DEAF : SJeanBoucheD.
- orso** : ANON. (1225*). *Orson de Beauvais*. Dees : Marne; compo. pic. mérid., ca 1225; ms. lorr., ca 1290; MS2 (ANON. 1899); DEAF : OrsonP.
- pap** : ANON. (1175*). *La vie du pape Saint Gregoire*. Ms. A1; Dees : nil; compo. Nord, ca 1175; ms. tour., ca 1238; MS1 (SOL 1977) DEAF : SGreg^a1S.
- pappreg2*** : SAINT GRÉGOIRE (OR.) (1180*). *Li dialogue Gregoire lo Pape, p. 5-52*. Dees : [Wallonie]; compo. lieg., ca 1180; ms. Est, ca 1210; MS3 (W. Foerster, Halle/Paris 1876, repr. 1965); DEAF : DialGregF.
- pen*** : ANDRIUS (1275*). *The penitance of Adam*. Dees : [Somme, Pas-de-Calais]; compo. nil, ca 1275; ms. pic., ca 1283; MS2 (E.C. Quinn, Mississippi, Romance Monographs 1980); DEAF : PenitAdam1/2Q.
- pera*** : CHRÉTIEN DE TROYES (1180*). *nil*. Dees : nil; compo. champ. merid., ca 1180; ms. champ., ca 1235; MS1 (transcr. inédite R.L.H. Lops); DEAF : PercLo_A.
- perb** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS B; Dees : nil; compo. champ. mérid., ca 1180; ms. bourg. sept., ca 1310; MS1 (transcr. inédite R. L. H. Lops); Dupliqué : percevalb.
- perc** : GERBERT DE MONTREUIL (1250*). *Continuation de Perceval*. Ms. A; Dees : nil; compo. pic., ca 1250; ms. pic. sept., ca 1275; MS1 (transcr. inédite R. L. H. Lops).
- percevalb** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS B; Dees : Haute-Marne; compo. champ. mérid., ca 1180; ms. bourg. sept., ca 1310; MS1 (transcr. inédite R. L. H. Lops); Dupliqué : perb.
- perf** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS F; Dees : nil; compo. champ. mérid., ca 1180; ms. Est, ca 1250; MS1 (transcr. inédite R. L. H. Lops).
- perh** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS H; Dees : nil; compo. champ. mérid., ca 1180; ms. a.n., ca 1350; MS1 (transcr. inédite R. L. H. Lops).
- perl** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS L; Dees : Aube; compo. champ. mérid., ca 1180; ms. champ. mérid., ca 1275; MS1 (transcr. inédite R. L. H. Lops).
- perm** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS M; Dees : Nièvre, Allier; compo. champ. mérid., ca 1180; ms. bourg., ca 1275; MS1 (transcr. inédite R. L. H. Lops).
- perp** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS P; Dees : Hainaut; compo. champ. mérid., ca 1180; ms. hain., ca 1250; MS1 (transcr. inédite R. L. H. Lops).
- perpraag** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS Prague; Dees : nil; compo. Troyes, ca 1180; ms. Nord Est, ca 1300; MS1 (transcr. inédite R. L. H. Lops).
- perq** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS Q; Dees : Haute-Marne; compo. champ. mérid., ca 1180; ms. champ. mérid., ca 1275; MS1 (transcr. inédite R. L. H. Lops).
- perr** : CHRÉTIEN DE TROYES (1180*). *Perceval*. Dees : [Hainaut]; compo. champ. mérid., ca 1180; ms. champ., ca 1237; MS1 (transcr. inédite R. L. H. Lops); inclus partiellement dans NCA_chret1.
- pers** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS S; Dees : Normandie; compo. champ. mérid., ca 1180; ms. traits norm., ca 1310; MS1 (transcr. inédite R. L. H. Lops).
- pert** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS T; Dees : nil; compo. champ. mérid., ca 1180; ms. pic. sept., ca 1275; MS1 (transcr. inédite R. L. H. Lops).
- peru** : CHRÉTIEN DE TROYES (1180*). *Perceval*. MS U; Dees : région parisienne; compo. champ. mérid., ca 1180; ms. Paris, ca 1330; MS1 (transcr. inédite R. L. H. Lops).

- plainte** : NICHOLAS BOZON (1312). *La plainte d'amour*. Dees : nil ; compo. a.n., 1312 ; ms. a.n., ca 1325 ; CR (BOZON 1905) ; DEAF : [PlainteAmV](#).
- poit** : ANON. (1217*). *Chronique traduite de l'Historia Karoli Magni et Rotholandi*. Dees : Vendée, Deux-Sèvres ; compo. saint., ca 1217 ; ms. saint., ca 1238 ; MS1 (AURACHER 1877) ; DEAF : [Turpin¹A](#).
- pon1** : ANON. (1213*). *La fille du comte de Pontieu*. Première main ; Dees : Somme, Pas-de-Calais ; compo. Nord, ca 1213 ; ms. Nord, ca 1290 ; MS2 (BRUNEL 1926).
- pon2** : ANON. (1238*). *La fille du comte de Pontieu*. Deuxième main ; Dees : Somme, Pas-de-Calais ; compo. Nord, ca 1238 ; ms. Nord, ca 1290 ; MS2 (BRUNEL 1926).
- pritheo** : ANON. (1290*). *Li priere Theophilus*. Dees : Somme, Pas-de-Calais ; compo. pic., ca 1290 ; ms. pic., ca 1300 ; CR2 (SCHELER 1877) ; DEAF : [PriereTheopS](#).
- prologueH*** : MARIE DE FRANCE (1165*). *nil*. Dees : nil ; compo. agn., ca 1165 ; ms. agn., ca 1275 ; MS (K. Warnke, Halle 1900) ; DEAF : [MarieProlW2](#).
- psautier** : ANON. (1125*). *Livre des psaumes*. Dees nil ; compo. a.n., ca 1125 ; ms. a.n., ca 1150 ; CR2 (MICHEL 1876), DEAF : [PsCambrM](#).
- pseuturp** : ANON. (1290*). *The burgundian translation of the Pseudo-Turpin Chronicle*. Dees : Haute-Saône ; compo. bourg., ca 1290 ; ms. *idem* ; MS (WALPOLE 1949, 1948) ; DEAF : [Turpin⁷W](#).
- queste** : ANON. (1228*). *La queste del saint Graal*. Ms. K, Dees : Nièvre, Allier ; compo. nil, ca 1228 ; ms. nil, ca 1250 ; CR3 (PAUPHILET 1923) ; DEAF : [SGraalivQuestP](#).
- raou*** : ANON. (1190*). *Raoul de Cambrai, v. 6250-8726 (seconde main)*. Dees : [Ardennes] ; compo. Nord-Est, ca 1190 ; ms. Nord, ca 1275 ; CR1 (P. Meyer et A. Longnon, SATF, Paris 1882) ; DEAF : [RCambr2M](#).
- reis** : ANON. (1175*). *Les quatre livres des Rois*. Ms. M ; Dees : nil ; compo. a.n., ca 1175 ; ms. a.n., ca 1190 ; MS (prob. CURTIUS 1911).
- remi*** : RICHIER (1283*). *La vie de saint Remi, poème du XIIIe s. par Richier. v. 1-3000*. Dees : [Ardennes] ; compo. champ. sept., ca 1283 ; ms. Ardennes, ca 1350 ; MS2 (W.N. Bolderston, London 1912) ; DEAF : [SRemiB](#).
- ren2*** : ANON. (1188*). *Le roman de Renart, branches II-VI, v. 3257-5550*. Dees : [Aube] ; compo. nil, ca 1188 ; ms. Est, ca 1213 ; MS (M. Roques, CFMA, Paris 1951) ; DEAF : [RenR](#).
- robert** : ANON. (1190*). *Robert le diable*. Dees : Somme, Pas-de-Calais ; compo. traits nomr. ?, ca 1190 ; ms. pic., ca 1275 ; CR (ANON. 1903) ; DEAF : [RobDiablE.L](#).
- robin** : ADAM DE LA HALLE (1225*). *Li Gieus de Robin et Marion*. Ms. P ; Dees : Somme ; compo. art., ca 1285 ; ms. Arras, 1295 ; MS (ADAM DE LA HALLE 1980).
- rolandox** : ANON. (1100*). *La Chanson de Roland*. Ms. O ; Dees : Angleterre ; compo. N.-O., ca 1100 ; ms. a.n., ca 1137 ; CR1 (SEGRE 1971) ; DEAF : [RoLS](#).
- roma*** : ANON. (1188*). [PVR] *Le Roman de Renart, br. VI, ms A E. Nieboer, Amsterdam (A) BN. f.fr. 20043 13e s.* Dees : [Haute-Marne] ; compo. nil, ca 1188 ; ms. pic., ca 1250 ; MS (transcr. inédite Nieboer) ; DEAF : [RenN_A](#).
- romb*** : ANON. (1188*). *Le roman de Renart, branche VI, ms. B*. Dees : [Haute-Marne] ; compo. nil, ca 1188 ; ms. Est, ca 1275 ; MS (transcr. inédite Nieboer) ; DEAF : [RenN_B](#).
- rombriv*** : ANON. (1188*). *nil*. Dees : nil ; compo. nil, ca 1188 ; ms. pic., ca 1290 ; MS (?) ; DEAF : [Ren_H](#).
- rombriva*** : ANON. (1188*). *nil*. Dees : nil ; compo. nil, ca 1188 ; ms. nil, nil ; MS (?) ; DEAF : [Ren_A](#).
- romc*** : ANON (1188*). *Le roman de Renart, branche VI, ms. C*. Dees : [Nievre, Allier] ; compo. nil, ca 1188 ; ms. nil, ca 1300 ; MS1 (transcr. inédite E. Nieboer) ; DEAF : [RenN_C](#).
- romd*** : ANON. (1188*). *Le roman de Renart, branche VI, ms. D*. Dees : [Region parisienne] ; compo. nil, ca 1188 ; ms. frc., 1339 ; MS1 (transcr. inédite E. Nieboer) ; DEAF : [RenN_D](#).
- rome*** : ANON. (1188*). *Le roman de Renart, branche VI, ms. E*. Dees : [Region parisienne] ; compo. nil, ca 1188 ; ms. nil, ca 1400 ; MS1 (transcr. inédite E. Nieboer) ; DEAF : [RenN_E](#).
- romh*** : ANON. (1188*). *Le roman de Renart, branche VI, ms. H*. Dees : [Ardennes] ; compo. nil, ca 1188 ; ms. pic., ca 1290 ; MS1 (transcr. inédite E. Nieboer) ; DEAF : [RenN_H](#).
- romi*** : ANON. (1188*). *nil*. Dees : nil ; compo. nil, ca 1188 ; ms. nil, nil ; MS1 (transcr. inédite E. Nieboer) ; DEAF : [RenN_I](#).
- romk*** : ANON. (1188*). [PVR] *Le Roman de Renart, br. VI, ms K E. Nieboer, Amsterdam Chantilly, Musee Conde, 472 13e ou 14e s.* Dees : [Somme, Pas-de-Calais] ; compo. nil, ca 1188 ; ms. hain., ca 1283 ; MS1 (transcr. inédite E. Nieboer) ; DEAF : [RenN_K](#).
- roml*** : ANON. (1188*). *Le roman de Renart, branche VI, ms. L*. Dees : [Haute-Marne] ; compo. nil, ca 1188 ; ms. nil, ca 1350 ; MS1 (transcr. inédite E. Nieboer) ; DEAF : [RenN_L](#).

- romm*** : ANON. (1188*). *Le roman de Renart, branche VI, ms. M.* Dees : [Nievre, Allier]; compo. nil, ca 1188; ms. nil, ca 1310; MS1 (transcr. inédite E. Nieboer); DEAF : RenN_M.
- romo*** : ANON. (1188*). *Le roman de Renart, branche VI, ms. O.* Dees : [Aube]; compo. nil, ca 1188; ms. nil, nil; MS1 (transcr. inédite E. Nieboer); DEAF : RenN_0.
- rose** : GUILLAUME DE LORRIS (1230*). *Le roman de la rose.* Dees : Nièvre, Allier; compo. traits orl., ca 1230; ms. orl., ca 1285; CR2 (GUILLAUME DE LORRIS 1934); DEAF : Rose_Lec.
- rou1*** : WACE (1167*). *nil.* Dees : nil; compo. norm., ca 1167; ms. frc., ca 1300; MS2 (A.J. Holden, Paris 1970); DEAF : RouH_CA.
- rou2*** : WACE (1167*). *nil.* Dees : nil; compo. norm., ca 1167; ms. frc., ca 1300; MS2 (A.J. Holden, Paris 1970); DEAF : RouH_2.
- rou3a*** : WACE (1167*). *Le roman de Rou de Wace, tome I, 3e partie v. 273-1018, ms. A.* Dees : [Angleterre]; compo. norm., ca 1167; ms. agn., ca 1210; CR2 (A.J. Holden, Paris 1970); DEAF : RouH_3a.
- rou3b*** : WACE (1167*). [PVR] *Le roman de Rou, de Wace, Chronique ascendante des Ducs de Normandie III, tome I, 3e partie vs 1019-1974* A.J. Holden, SATF, Paris 1970-73, 3 volumes A, Brit.Mus. Roy. 4.C.XI deb.13e s. Dees : [Angleterre]; compo. norm., ca 1167; ms. agn., ca 1210; CR2 (A.J. Holden, Paris 1970); DEAF : RouH_3b.
- sage*** : ANON. (1250*). *Les sept sages de Rome.* Dees : [Region parisienne]; compo. Traits de l'Ouest., ca 1250; ms. frc., ca 1290; MS1 (C.R.A.L. (Derniame/Henin/Nais), Nancy 1981); DEAF : SSagAD.
- sapient*** : SAINT GRÉGOIRE (1190*). *Sermo de sapientia, dans : Li dialogue Gregoire lo pape.* Dees : [Wallonie]; compo. Liege, ca 1190; ms. Est, ca 1210; MS2 (W. Foerster, Halle/Paris 1876, p. 283-298); DEAF : DialGregF.
- sept*** : ANON. (1183*). *Le roman des sept sages, ms. K, v. 1-810.* Dees : [Somme, Pas-de-Calais]; compo. Sud-Ouest, ca 1183; ms. pic., 1285; MS2 (J. Misrahi, Paris 1933); DEAF : SSagOctM.
- sergbH** : GUISCHART DE BEAULIU (ca 1190). *Sermon.* Compo. a.n., ca 1190; ms. a.n., ca 1250; CR1 (BEAULIU 1909), DEAF : GuischartG.
- sergb0** : GUISCHART DE BEAULIU (1190*). *Sermon.* Ms. O; Dees : nil; compo. a.n., ca 1190; ms. a.n., ca 1277; MS (GUISCHART DE BEAULIU 1909); DEAF : GuischartG.
- songe*** : RAOUL DE HOUDENC (1214*). *Le songe d'enfer de Raoul de Houdenc.* Dees : [Yonne]; compo. Beauvais, ca 1214; ms. nil, ca 1288; CR3 (Ph. Lebesgue, Paris 1908); DEAF : SongeEnfL.
- songe14*** : PHILIPPE DE MÉZIÈRES (1389). *nil.* Dees : nil; compo. pic., 1389; ms. nil, ca 1400; MS1 (?); DEAF : PhMézPeL_A.
- stsilv*** : ANON. (1200*). *La vie de Saint-Silvestre.* Dees : [Indre-et-Loire]; compo. Ouest, ca 1200; ms. tour., ca 1210; MS2 (A. Planchenault?); DEAF : SSilvCroixP.
- sully2** : MAURICE DE SULLY (ca 1250). *Sermons de Maurice de Sully.* Compo. La Rochelle, ca 1250; ms. poit., ca 1250; MS2 (BOUCHERIE 1873), DEAF : SermMaurpB.
- teo2** : ADGAR (1150*). *Le dit de Théophile.* Ms. M; Dees : nil; compo. a.n., ca 1150; ms. a.n., ca 1210; MS2 (WEBER 1877).
- thA*** : ANON. (1160*). [PVR] *Le roman de Thebes, fragments d'Angers L. Constans, SATF, Paris 1890, vol.II, appendice VI (D) fin 12e s.* Dees : [Vendee, Deux-Sevres]; compo. poit., ca 1160; ms. poit., ca 1200; MS (L. Constans, SATF, Paris 1890); DEAF : ThebesC_D.
- the*** : RUTEBEUF (1261*). *Le miracle de Théophile de Rutebeuf.* Dees : [Region parisienne]; compo. champ. merid., ca 1261; ms. frc., ca 1288; nil (G. Frank, CFMA, Paris 1949); DEAF : RutebTheophF.
- thebe*** : ANON. (1250*). *Le roman de Thèbes, Tome 1, ms. C, v. 1-766.* Dees : [Region parisienne]; compo. poit., ca 1250; ms. frc., ca 1250; MS2 (G. Raynaud de Lage, 2 vol., CFMA, Paris 1966); DEAF : ThebesR_C.
- thebefrag*** : ANON. (1160*). *Les fragments d'Angers du Roman de Thèbes, ms. D.* Dees : [Vendee, Deux-Sevres]; compo. poit., ca 1160; ms. poit., ca 1200; CR2 (G. Raynaud de Lage, Romania 90 1969, p. 403-409); DEAF : ThebesR_D.
- thibo*** : GUILLAUME D'OYE (1290*). *La vie de saint Thibaut.* Dees : [Haute-Marne]; compo. bourg., ca 1290; ms. bourg., ca 1290; MS2 (H.E. Manning, New York 1929); DEAF : SThibA1M.
- trib*** : BÉROUL (1188*). [PVR] *Le roman de Tristan de Beroul, V.1-1007* E. Muret, CFMA, 4me edition, Paris 1962 BN ffr 2171 2e moitié 2 13e s. Dees : [Nievre, Allier]; compo. Ouest, ca 1188; ms. nil, ca 1275; MS1 (E. Muret, CFMA, Paris 1962); DEAF : TristBérM5*.
- troi*** : ANON. (1288*). *Le roman de Troie en prose, par. 1-19.* Dees : [Marne]; compo. champ., ca 1288; ms. Marne, ca 1288; MS1 (Vieillard, Cologny-Genève 1979); DEAF : TroiePr15V.

- troifr*** : BENOÎT DE SAINTE-MAURE (1170*). [PVR] *Fragments du Roman de Troie P. Meyer, Rom. XVIII (1889), p.70-106 Bibl. de l'Universite, fragments de Bale, p.73-87 fin 12e s.* Dees : [Angleterre]; compo. poit., ca 1170; ms. Angleterre, ca 1190; CR1 (P. Meyer, Rom. XVIII 1889, p.70-106); DEAF : BenTroieM*.
- turp** : ANON. (1217*). *Chronique dite Saintongeaise.* Dees : [Vendée, Deux-Sèvres]; compo. saint., ca 1217; ms. poit, ca 1275; MS (MANDACH 1970); DEAF : Turpin¹M.
- vache*** : DROUART LA VACHE (1290). *Li livres d'amours de Drouart La Vache, v. 1-3010.* Dees : [Aisne]; compo. champ., 1290; ms. Aisne, ca 1300; MS1 (R. Bossuat, Paris 1926); DEAF : DrouartB.
- vcou*** : ANON. (1250*). *Li ver de Couloigne.* Dees : [Somme, Pas-de-Calais]; compo. pic. orient., ca 1250; ms. art., ca 1290; CR1 (H. Kleineidam, Munchen 1968); DEAF : VerCou¹K.
- vergia** : ANON. (1240*). *La chastelaine de Vergi.* Ms. A; Dees : Oise; compo. norm., ca 1240; ms. pic., 1289; MS1 (STUIP 1970); DEAF : ChastVergiS.
- vergib** : ANON. (1240*). *La chastelaine de Vergi.* Ms. B; Dees : Normandie; compo. norm., ca 1240; ms. norm., ca 1300; MS1 (STUIP 1970); DEAF : ChastVergiS.
- vergic** : ANON. (1240*). *La chastelaine de Vergi.* Ms. C; Dees : Yonne; compo. norm., ca 1240; ms. frc., ca 1288; MS1 (STUIP 1970); DEAF : ChastVergiS.
- vergie** : ANON. (1240*). *La chastelaine de Vergi.* Ms. E; Dees : Aube; compo. norm., ca 1240; ms. champ. mérid., ca 1310; MS1 (STUIP 1970); DEAF : ChastVergiS.
- vergif** : ANON. (1240*). *La chastelaine de Vergi.* Ms. F; Dees : Normandie; compo. norm., ca 1240; ms. pic., ca 1318; MS1 (STUIP 1970); DEAF : ChastVergiS.
- vergig** : ANON. (1240*). *La chastelaine de Vergi.* Ms. G; Dees : Aisne; compo. norm., ca 1240; ms. nil, ca 1290; MS1 (STUIP 1970); DEAF : ChastVergiS.
- vergih** : ANON. (1240*). *La chastelaine de Vergi.* Ms. H; Dees : Haute-Marne; compo. norm., ca 1240; ms. nil, ca 1310; MS1 (STUIP 1970); DEAF : ChastVergiS.
- vergii** : ANON. (1240*). *La chastelaine de Vergi.* Ms. I; Dees : Normandie; compo. norm., ca 1240; ms. norm., ca 1300; MS1 (STUIP 1970); DEAF : ChastVergiS.
- vergik** : ANON. (1240*). *La chastelaine de Vergi.* Ms. B; Dees : Aisne; compo. norm., ca 1240; ms. pic., ca 1350; MS1 (STUIP 1970); DEAF : ChastVergiS.
- vergil** : ANON. (1240*). *La chastelaine de Vergi.* Ms. L; Dees : région parisienne; compo. norm., ca 1240; ms. frc., ca 1325; MS1 (STUIP 1970); DEAF : ChastVergiS.
- verite*** : ANON. (nil). [PVR] *le roman de Verite et de Faussete Leuven BN Moreau 1727 15e s.* Dees : nil; compo. nil, nil; ms. nil, ca 1750; MS (M. Goussaert, Louvain 1978); DEAF : RomVeritG*.
- verson*** : ESTOUT DE GOZ (1247). *Le conte des vilains de Verson par Estout de Gox.* Dees : [Normandie]; compo. norm., 1247; ms. nil, ca 1250; MS2 (L. Delisle, Etudes sur la condition de la classe agricole, Paris 1903); DEAF : VilVersonD.
- vilea*** : GEOFFROY DE VILLEHARDOIN (1205*). *La conquête de Constantinople de Villehardouin, ms. O.* Dees : [Bourgogne]; compo. champ., ca 1205; ms. nil, ca 1375; CR2 (E. Faral, Paris 1961); DEAF : VillehF.
- vilhar*** : GEOFFROY DE VILLEHARDOIN (1205*). *La conquête de Costentinoble de Josfroi de Viieharduyn, ms. B.* Dees : [Aisne]; compo. champ., ca 1205; ms. frc., ca 1290; MS1 (C.R.A.L., Nancy 1978); DEAF : VillehC*.
- volu*** : OMONT (1265). *Li volucraires.* Dees : [Ardennes]; compo. champ., 1265; ms. champ., ca 1288; MS2 (H. Kleineidam, ZfRP 1970, p. 10-21); DEAF : VolucrK.
- wallo** : ANON. (1290*). *Gloses wallonnes.* Dees : Wallonie; compo. wall., ca 1290; ms. wall., ca 1290; MS (WILMOTTE 1891).
- wita*** : ANON. (1238*). *Li romans de Witasse le moine, v. 1-853.* Dees : [Somme, Pas-de-Calais]; compo. pic., ca 1238; ms. pic., 1285; MS3 (D.J. Conlon, Chapel Hill, 1972); DEAF : EustMoineC.
- yonecH** : MARIE DE FRANCE (1165*). *Lai de Yonec.* Ms. H; Dees : nil; compo. a.n., ca 1165; ms. a.n., ca 1275; MS (éd. non identifiée).
- yonecP** : MARIE DE FRANCE (1165*). *Lai de Yonec.* Ms. P; Dees : nil; compo. a.n., ca 1165; ms. pic., ca 1290; MS (éd. non identifiée).
- yonecQ** : MARIE DE FRANCE (1165*). *Lai de Yonec.* Ms. Q; Dees : nil; compo. a.n., ca 1165; ms. frc., ca 1339; MS (éd. non identifiée).
- yonecS** : MARIE DE FRANCE (1165*). *Lai de Yonec.* Ms. S; Dees : nil; compo. a.n., ca 1165; ms. frc., ca 1300; MS (éd. non identifiée).

- yva*** : CHRÉTIEN DE TROYES (1177*). [PVR] *Le chevalier au lion de Chretien de Troyes v.1-1000 R.L.H. Lops (A) Chantilly 472 fin 13e s.* Dees : [Haute-Marne]; compo. champ. merid., ca 1177; ms. hain., ca 1283; MS1 (transcr. inédite R.L.H. Lops); DEAF : YvainL_A.
- yvf*** : CHRÉTIEN DE TROYES (1177*). *nil.* Dees : nil; compo. champ. merid., ca 1177; ms. pic., ca 1237; MS1 (transcr. inédite R.L.H. Lops); DEAF : YvainL_F.
- yvg*** : CHRÉTIEN DE TROYES (1177*). *Le chevalier au lion de Chrétien de Troyes, ms. G, v. 1-1000.* Dees : [Aube]; compo. champ. merid., ca 1177; ms. champ., ca 1250; MS1 (transcr. inédite R.L.H. Lops); DEAF : YvainL_G.
- yvh** : CHRÉTIEN DE TROYES (1177*). *Le Chevalier au lion.* Ms. H; Dees : nil; compo. champ. mérid., ca 1177; ms. champ., ca 1235; MS (transcr. inédite R. L. H. Lops); également contenu dans chret2.
- yvp*** : CHRÉTIEN DE TROYES (1177*). *Le chevalier au lion de Chrétien de Troyes, ms. P, v. 1-1000.* Dees : [Somme, Pas-de-Calais]; compo. champ. merid., ca 1177; ms. pic., ca 1317; MS (transcr. inédite R.L.H. Lops); DEAF : YvainL_P.
- yvs*** : CHRÉTIEN DE TROYES (1177*). *Le chevalier au lion de Chrétien de Troyes, ms. S, v. 1-1000.* Dees : [Somme, Pas-de-Calais]; compo. champ. merid., ca 1177; ms. nil, ca 1300; MS1 (transcr. inédite R.L.H. Lops); DEAF : YvainL_S.
- yvv*** : CHRÉTIEN DE TROYES (1177*). *Le chevalier au lion de Chrétien de Troyes v. 1-1000, ms. V.* Dees : [Nievre, Allier]; compo. champ. merid., ca 1177; ms. frc., ca 1290; MS1 (transcr. inédite R.L.H. Lops); DEAF : YvainL_V.
- yzop** : ANON. (1275*). *Ysopet de Lyon.* Dees : Franche-Comté; compo. Est, ca 1275; ms. frcomt., ca 1290; MS (ANON. 1882); DEAF : YsLyonF.

C.2 BFM : liste des textes cités

Les localisations et dates viennent des métadonnées de la BFM. L'édition est indiquée à la fin de la citation, par nom de l'éditeur le cas où l'auteur est anonyme. La liste donnée se restreint aux textes cités dans ce travail.

- AlexisRaM** : ANON. (1050*). *Vie de saint Alexis.* Compo. ca 1050, ms. ca 1120, dial. norm. (RAINSFORD et MARCHELLO-NIZIA 2018).
- aucassin** : ANON. (1208*). *Aucassin et Nicolette.* Compo. dern. q. xii^e s. ou 1^{ère} m. xiii^e s., ms. fin xiii^e s., dial. pic. (ROQUES 1929a).
- becket** : GUERNES DE PONT-SAINTE-MAXENCE (ca 1173*). *Vie de saint Thomas Becket.* Compo. entre 1172 et 1174, ms. déb. xiii^e s., dial. Ouest (GUERNES DE PONT-SAINTE-MAXENCE 1936).
- ErecKu** : CHRÉTIEN DE TROYES (1170*). *Erec et Enide.* Compo. ca 1170, ms. ca 1235, dial. champ. (CHRÉTIEN DE TROYES 2009b).
- eulaliBfm** : ANON (881*). *Séquence de sainte Eulalie.* Compo. peu après 881, ms. ca 900, (LAVRENTIEV, GUILLOT-BARBANCE et RAINSFORD 2018).
- jonas** : ANON. (ca 945*). *Sermon sur Jonas.* Compo. prob. entre 938 et 952, ms. 2^e q. x^e s., dial. fla. (POERCK 1956).
- Lapidal** : PHILIPPE DE THAON (ca 1133*). *Lapidaire alphabétique.* Compo. 1^{er} t. xii^e s., ms. ca 1200, dial. a.n. (THAON 1924).
- oxfps** : ANON. (ca 1125*). *Psautier d'Oxford.* Compo. 1^{ère} m. xii^e s., ms. m. xii^e s., dial. a.n. (MICHEL 1860).
- passion** : ANON. (1000*). *Passion de Jésus-Christ ou Passion de Clermont.* Compo. et ms. ca 1000, dial. fr.-oc. (AVALLE 1962).
- PercevalKu** : CHRÉTIEN DE TROYES (1183*). *Conte du Graal (Perceval).* Compo. ca 1181-1185, ms. ca 1235, dial. champ. (CHRÉTIEN DE TROYES 2009a).
- roland** : ANON. (ca 1100). *Chanson de Roland.* Compo. ca 1100, ms. ca 1120, dial. norm. (MOIGNET 1972a).
- slethgier** : ANON. (1000). *Vie de saint Léger.* Compo. ca 1000, ms. ca 1000, dial. fr.-oc. (AVALLE 1967).
- strasBfm** : ANON (842*). *Serments de Strasbourg.* Compo. ca 842, ms. ca 1000, (LAVRENTIEV, RAINSFORD et GUILLOT-BARBANCE 2014).

D

Corpus manuscrits : liste des témoins

D.1 Manuscrits des trouvères

- a** : *Chansonnier français a* (BAV Reg. lat. 1490). URL : https://digi.vatlib.it/view/MSS_Reg.lat.1490.
- i** : *Miracles de Notre Dame, en vers, etc.* (BNF fr 12483). URL : <http://gallica.bnf.fr/ark:/12148/btv1b8454680s>.
- K** : *Recueil de chansons du XIII^e siècle, avec musique notée* (BNF fr 5198). URL : <http://gallica.bnf.fr/ark:/12148/btv1b550063912>.
- M** : *Français 844* (BNF fr 844). URL : <https://gallica.bnf.fr/ark:/12148/btv1b84192440>.
- N** : *Français 845* (BNF fr 845). URL : <http://gallica.bnf.fr/ark:/12148/btv1b6000955r>.
- O** : *Chansonnier, dit Chansonnier Cangé* (BNF fr 846). URL : <https://gallica.bnf.fr/ark:/12148/btv1b6000950p>.
- P** : *Français 847* (BNF fr 847). URL : <http://gallica.bnf.fr/ark:/12148/btv1b8454673n>.
- U** : *Chansonnier français, dit de Saint-Germain-des-Prés* (BNF fr 20050). URL : <http://gallica.bnf.fr/ark:/12148/btv1b60009580>.
- V** : *Chansonnier français — Traité des quatre nécessaires — Richard de Fournival, Bestiaire d'Amour — Chansons à la Vierge* (BNF fr 24406). URL : <http://gallica.bnf.fr/ark:/12148/btv1b84386028>.
- X** : *Chansons françaises, avec musique notée (Chansonnier Clairambault)* (BNF fr 1050). URL : <http://gallica.bnf.fr/ark:/12148/btv1b530003205>.

D.2 Manuscrits d'*Aucassin et Nicolette* et du *Gieus de Robin et de Marion*

- Jean Bodel, Le Vilain de Farbu, etc.* (BNF fr 2168). URL : <https://gallica.bnf.fr/ark:/12148/btv1b525135854>.
- Manuscrit de la Vallière* (BNF fr 25566). URL : <https://gallica.bnf.fr/ark:/12148/btv1b6001348v>.

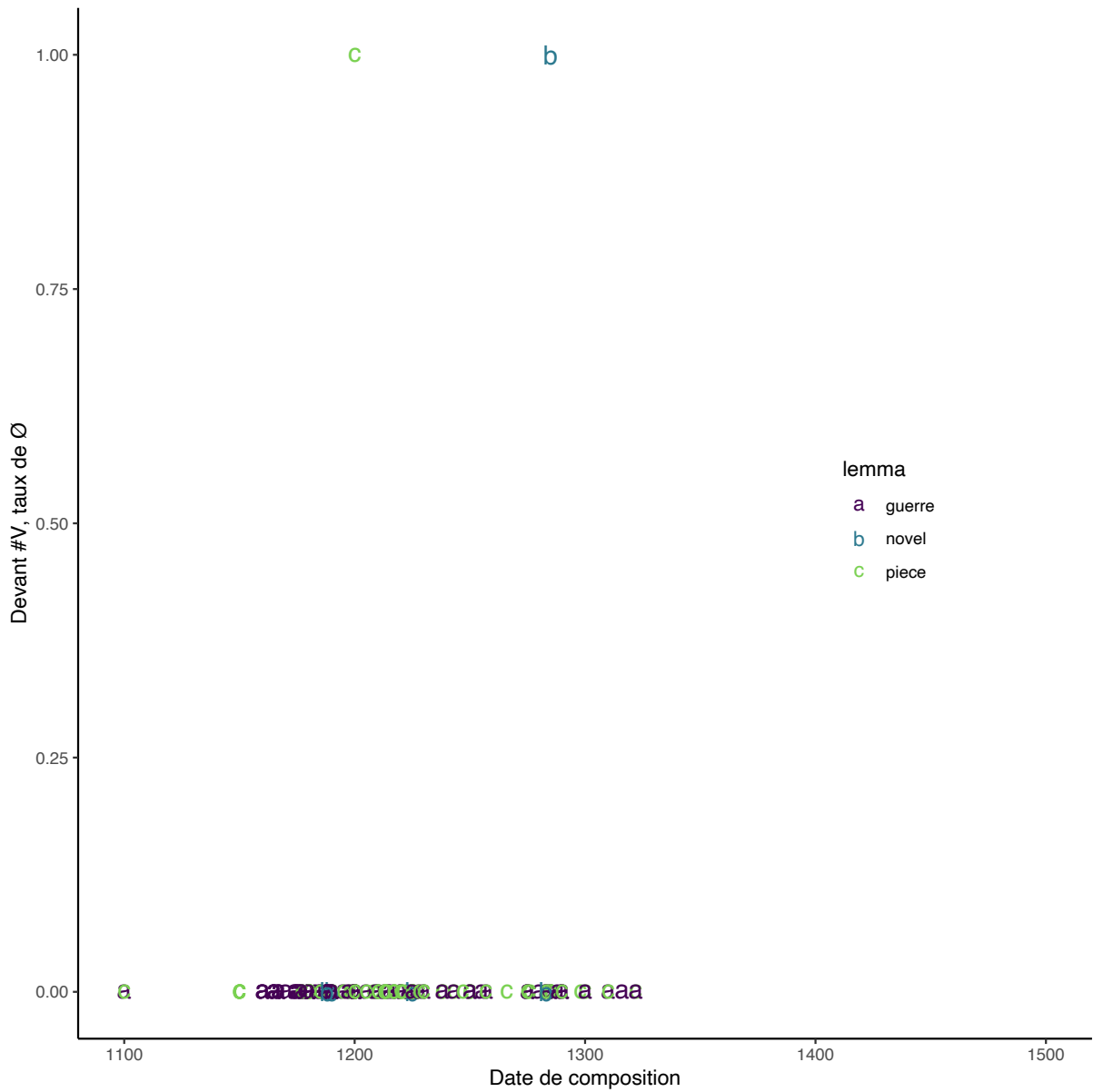
Figures rejetées en annexes

```

1 <s n="nca:355312_rolandox:1">
2 <w id="w_rolandox_s355312_1"><txm:form><txm:form>carles</txm:form><txm:ana resp="#src"
  type="#src">--</txm:ana><txm:ana resp="#src" type="#taggerpos">NPR</txm:ana><txm:ana
  resp="#src" type="#pos">NPR_suj_masc_sg</txm:ana><txm:ana resp="#src" type="#lemma">
  charle</txm:ana><txm:ana resp="#src" type="#deespos">011</txm:ana></txm:form><txm:ana
  resp="none" type="#type">w</txm:ana><txm:ana resp="none" type="#n">6</txm:ana></w>
3 <w id="w_rolandox_s355312_2"><txm:form><txm:form>li</txm:form><txm:ana resp="#src" type="
  #src">--</txm:ana><txm:ana resp="#src" type="#taggerpos">DET_def</txm:ana><txm:ana
  resp="#src" type="#pos">DET_def_suj_masc_sg</txm:ana><txm:ana resp="#src" type="#
  lemma">le</txm:ana><txm:ana resp="#src" type="#deespos">101</txm:ana></txm:form><txm:
  ana resp="none" type="#type">w</txm:ana><txm:ana resp="none" type="#n">7</txm:ana></w>
4 <w id="w_rolandox_s355312_3"><txm:form><txm:form>reis</txm:form><txm:ana resp="#src" type
  ="#src">--</txm:ana><txm:ana resp="#src" type="#taggerpos">NOM</txm:ana><txm:ana resp
  ="#src" type="#pos">NOM_suj_masc_sg</txm:ana><txm:ana resp="#src" type="#lemma">roi</
  txm:ana><txm:ana resp="#src" type="#deespos">001</txm:ana></txm:form><txm:ana resp="
  none" type="#type">w</txm:ana><txm:ana resp="none" type="#n">8</txm:ana></w>
5 <w id="w_rolandox_s355312_4"><txm:form><txm:form>nostre</txm:form><txm:ana resp="#src"
  type="#src">--</txm:ana><txm:ana resp="#src" type="#taggerpos">DET_poss</txm:ana><txm:
  ana resp="#src" type="#pos">DET_poss_suj_masc_sg</txm:ana><txm:ana resp="#src" type="
  #lemma">nostre</txm:ana><txm:ana resp="#src" type="#deespos">161</txm:ana></txm:form
  ><txm:ana resp="none" type="#type">w</txm:ana><txm:ana resp="none" type="#n">9</txm:
  ana></w>
6 <w id="w_rolandox_s355312_5"><txm:form><txm:form>empere</txm:form><txm:ana resp="#src"
  type="#src">--</txm:ana><txm:ana resp="#src" type="#taggerpos">NOM</txm:ana><txm:ana
  resp="#src" type="#pos">NOM_suj_masc_sg</txm:ana><txm:ana resp="#src" type="#lemma">
  empire</txm:ana><txm:ana resp="#src" type="#deespos">001</txm:ana></txm:form><txm:ana
  resp="none" type="#type">w</txm:ana><txm:ana resp="none" type="#n">10</txm:ana></w>
7 <w id="w_rolandox_s355312_6"><txm:form><txm:form>magnes</txm:form><txm:ana resp="#src"
  type="#src">--</txm:ana><txm:ana resp="#src" type="#taggerpos">ADJ</txm:ana><txm:ana
  resp="#src" type="#pos">ADJ_suj_masc_sg</txm:ana><txm:ana resp="#src" type="#lemma">
  maigne</txm:ana><txm:ana resp="#src" type="#deespos">021</txm:ana></txm:form><txm:ana
  resp="none" type="#type">w</txm:ana><txm:ana resp="none" type="#n">11</txm:ana></w>
8 </s>

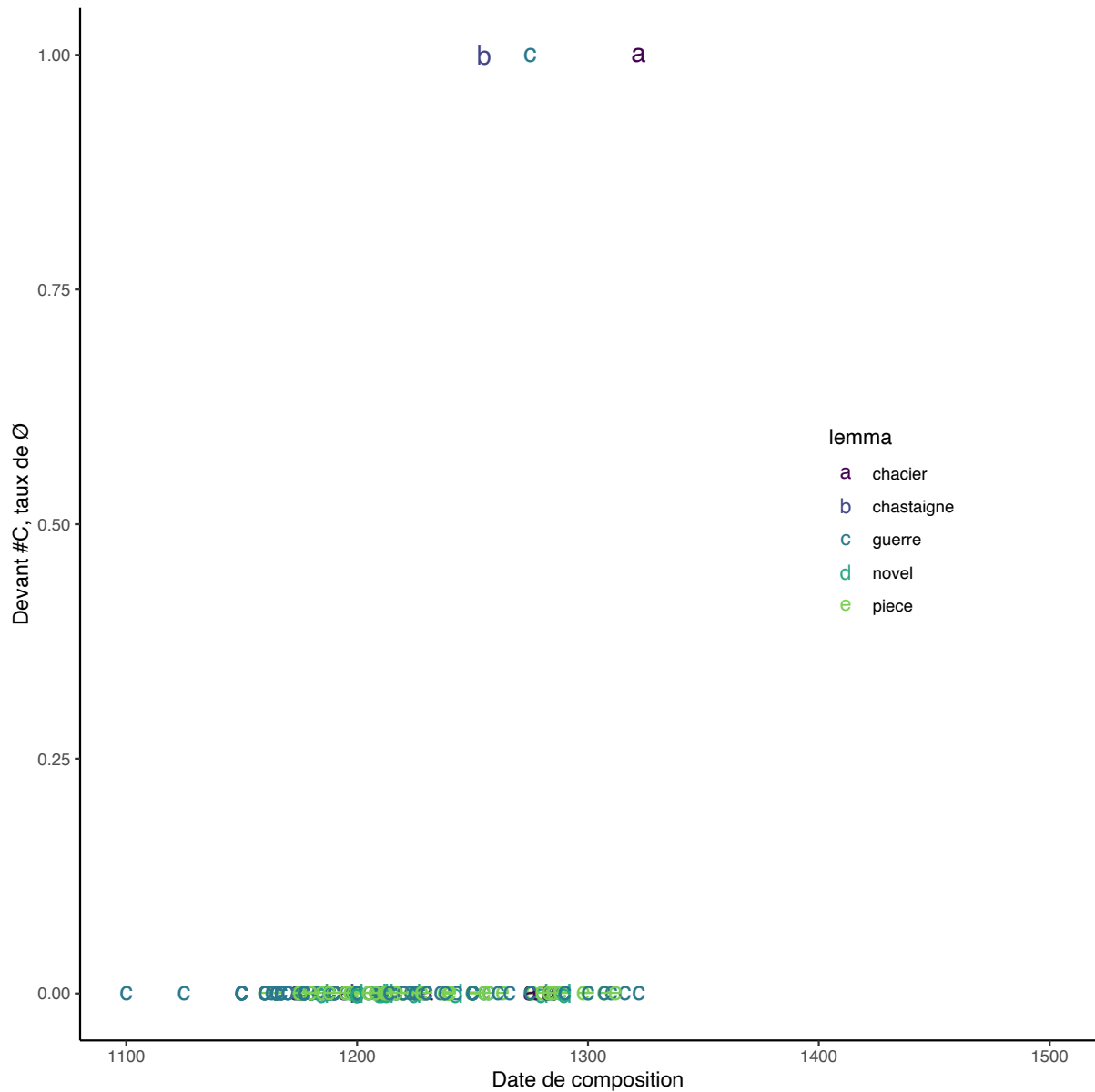
```

FIGURE E.1 – *Chanson de Roland*, v. 2, dans le format du fichier rolandox.xml du NCA importé dans TXM.



lemma	nb.e	nb.Ø	tot.	tx.Ø
novel	5	1	6	0.17
piece	149	1	150	0.01
guerre	165	0	165	0.00

FIGURE E.2 – Distribution diachronique des variantes de noms en défaut de <e> devant initiale vocalique, en fonction des lemmes.



lemma	nb.e	nb.Ø	tot.	tx.Ø
chastaigne	1	1	2	0.50
chacier	5	1	6	0.17
guerre	362	2	364	0.01
piece	181	0	181	0.00
novel	10	0	10	0.00

FIGURE E.3 – Distribution diachronique des variantes de noms en défaut de <e> devant initiale consonantique, en fonction des lemmes.

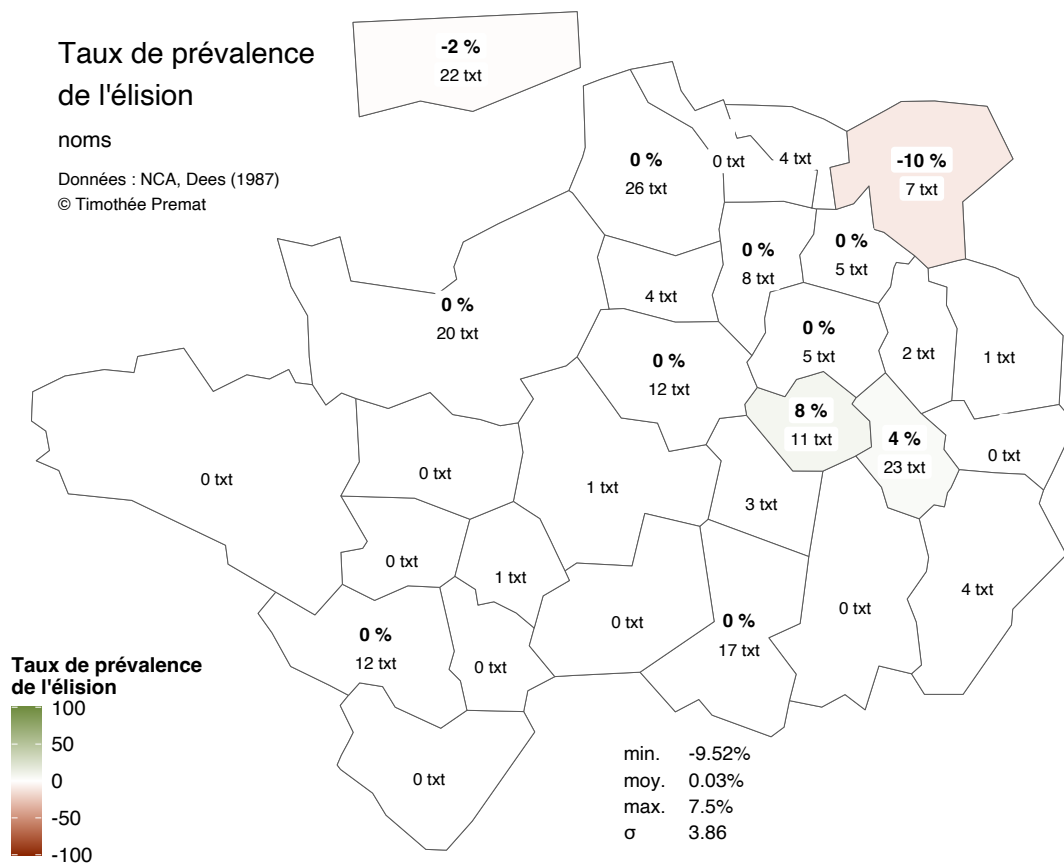


FIGURE E.4 – Distribution diatopique de la prévalence de l'élision, noms.

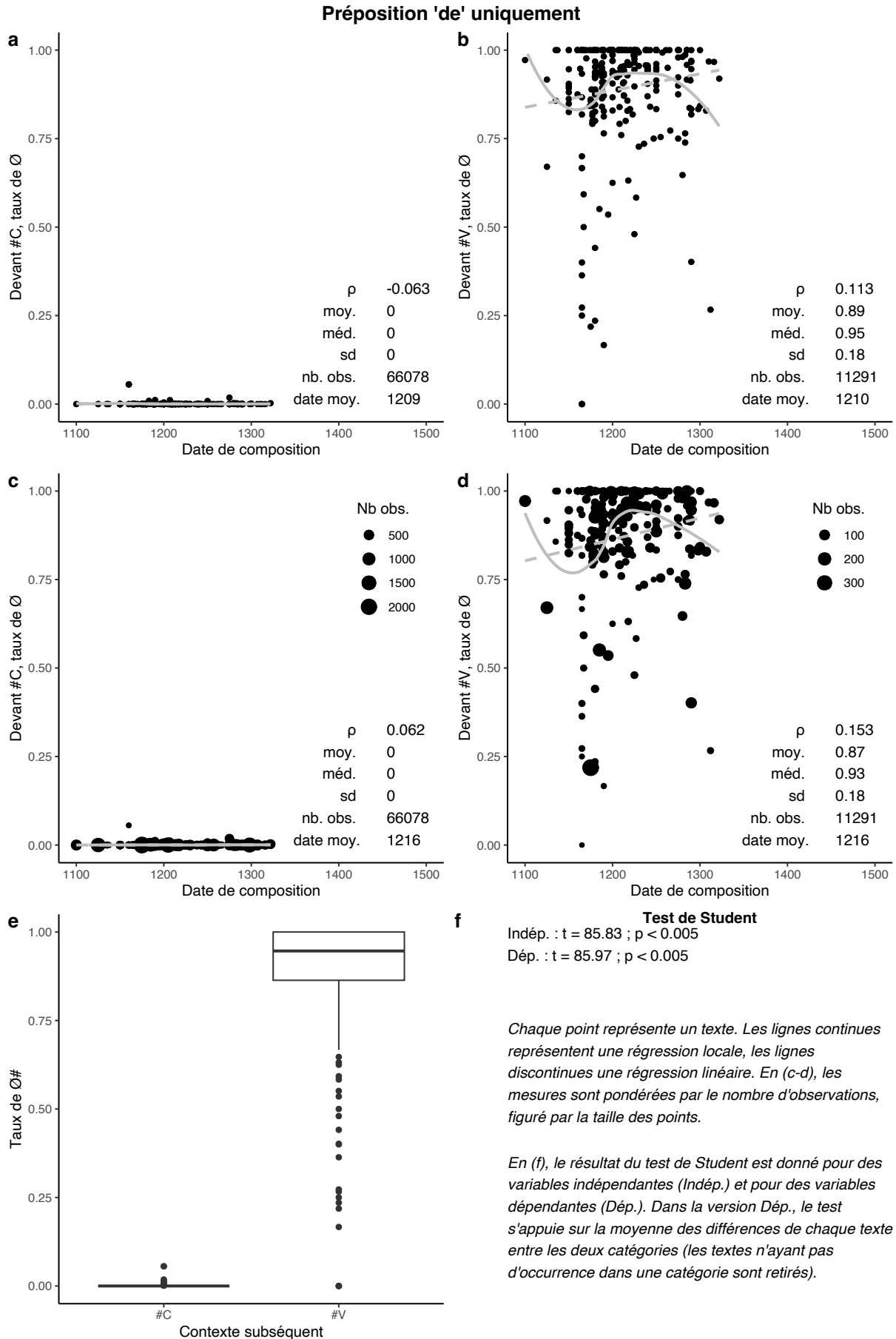


FIGURE E.5 – Distribution contextuelle des variantes de prépositions « de » en défaut de <e>.

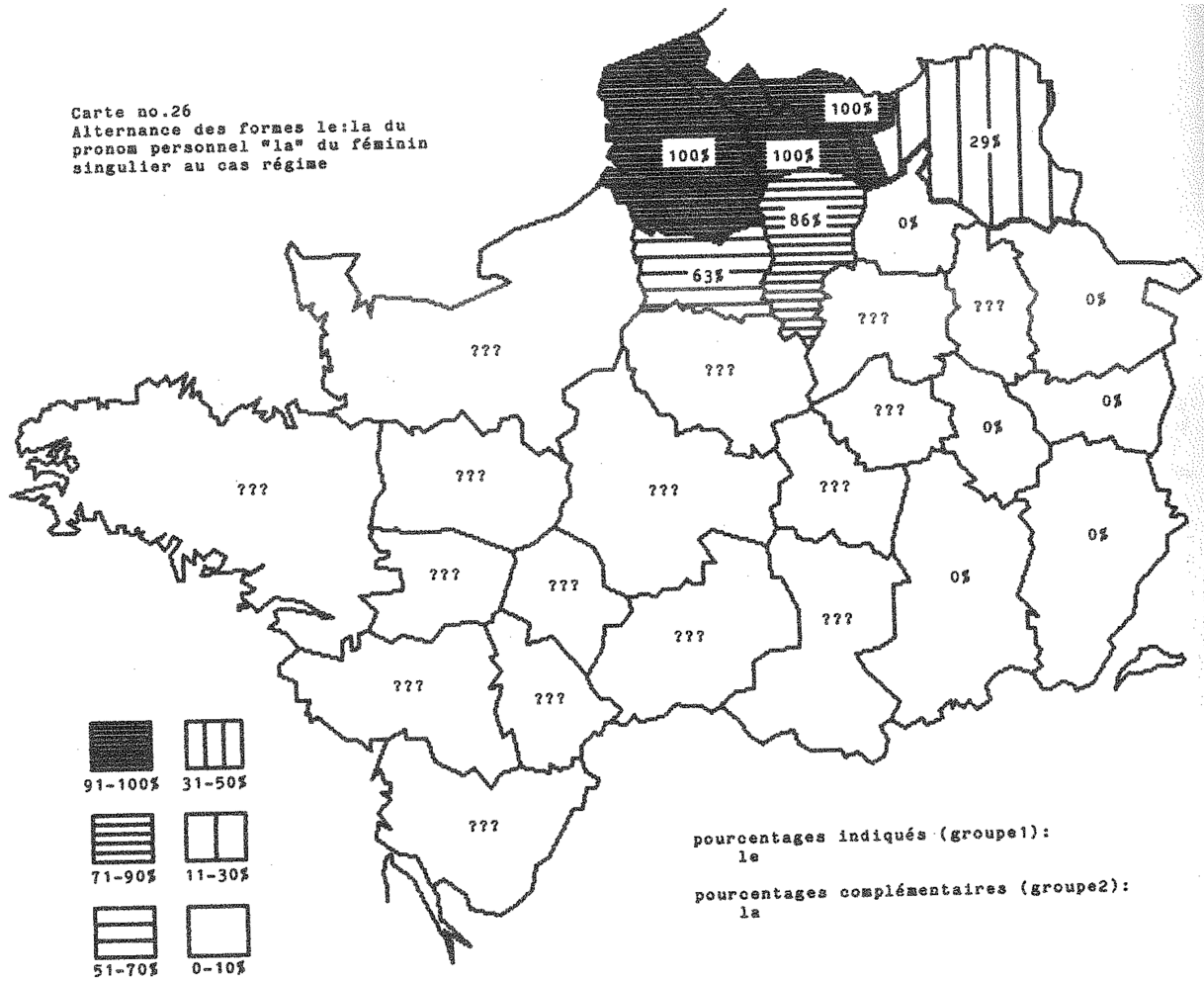


FIGURE E.6 – Distribution diatopique des formes picardes *le* du pronom personnel de 3sg CR FÉM (*vs la*); DEES 1980, p. 28, c. 26.

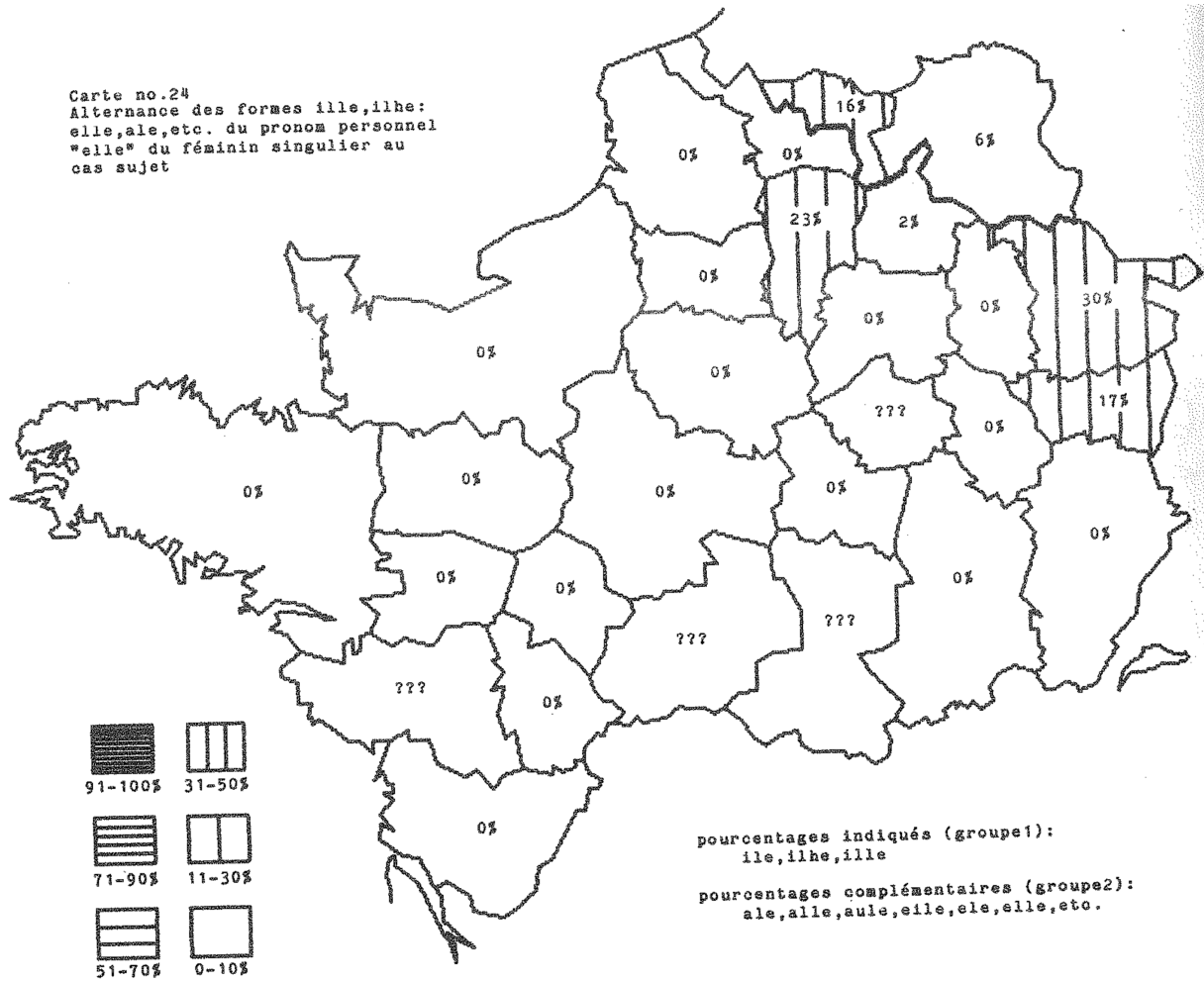
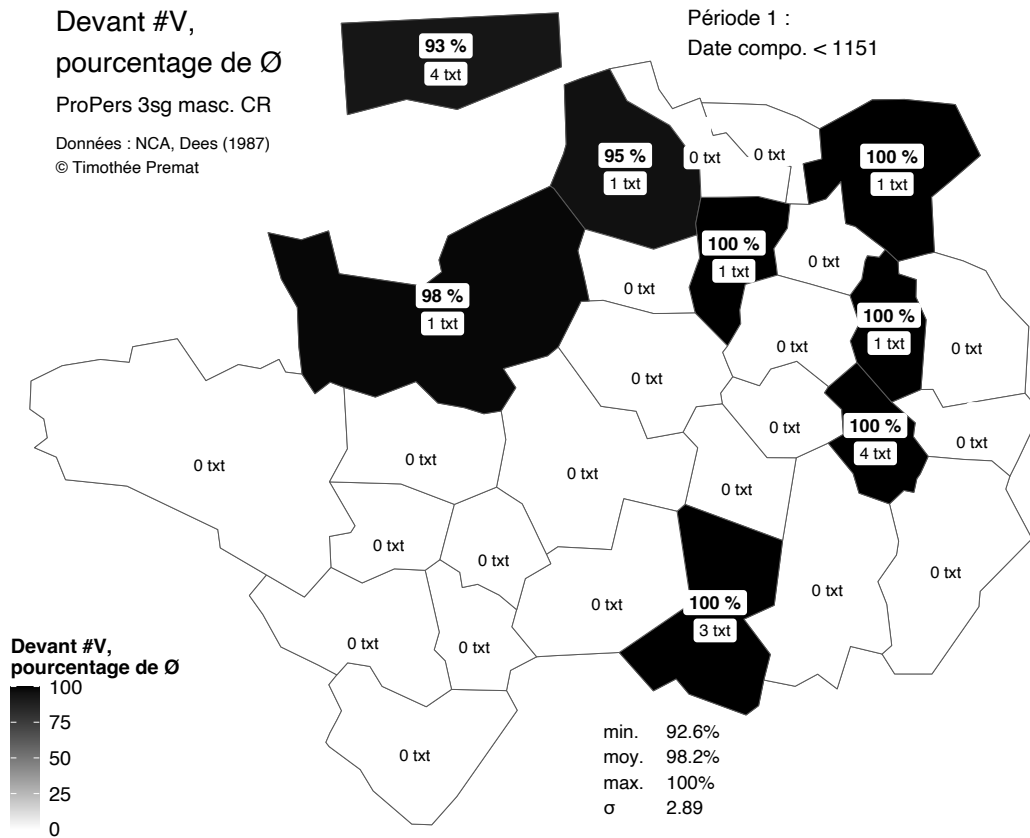


FIGURE E.7 – Distribution diatopique des formes en <i> initial du pronom personnel de 3sg CS FÉM; DEES 1980, p. 26, c. 24.

a



b

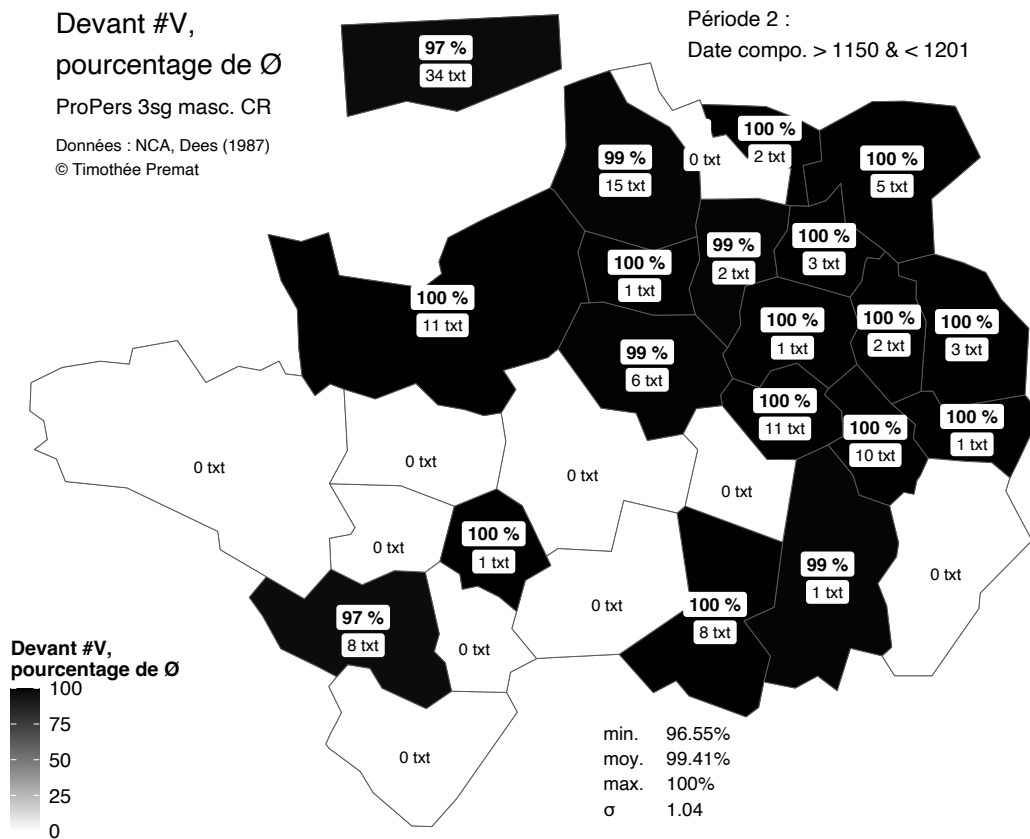
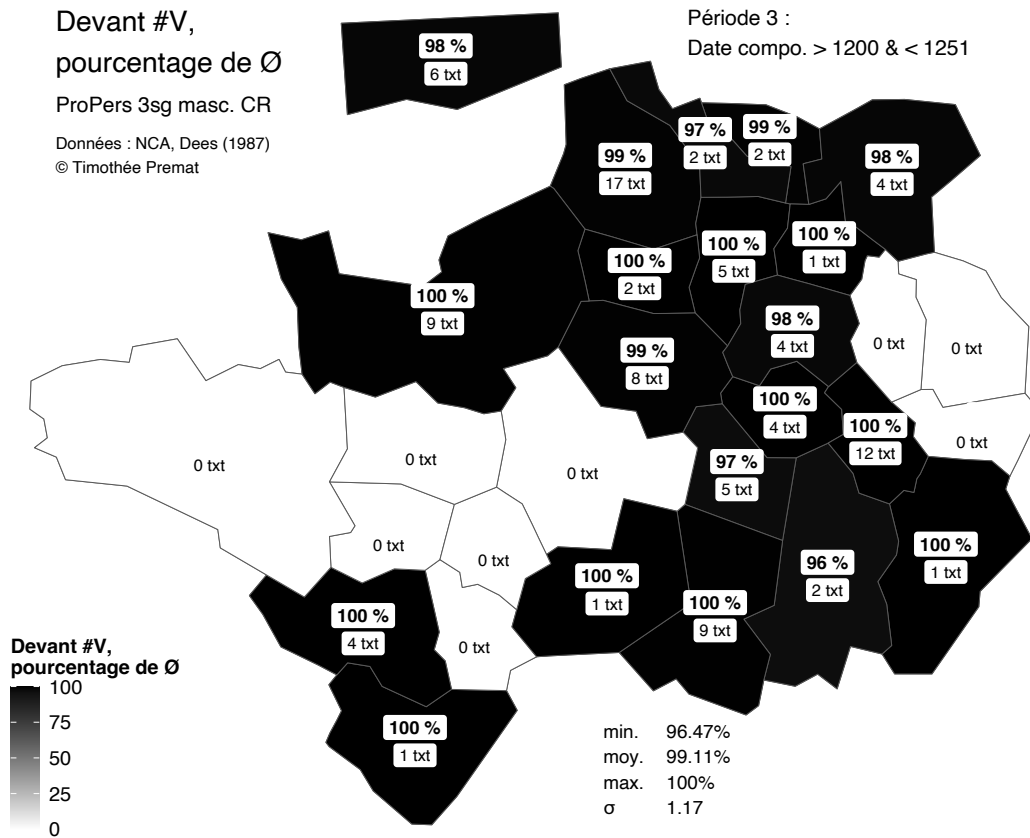


FIGURE E.8 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1050) et 2 (b : 1151-1200), pronoms personnels masculins régimes de troisième personne du singulier, devant initiale vocalique.

a



b

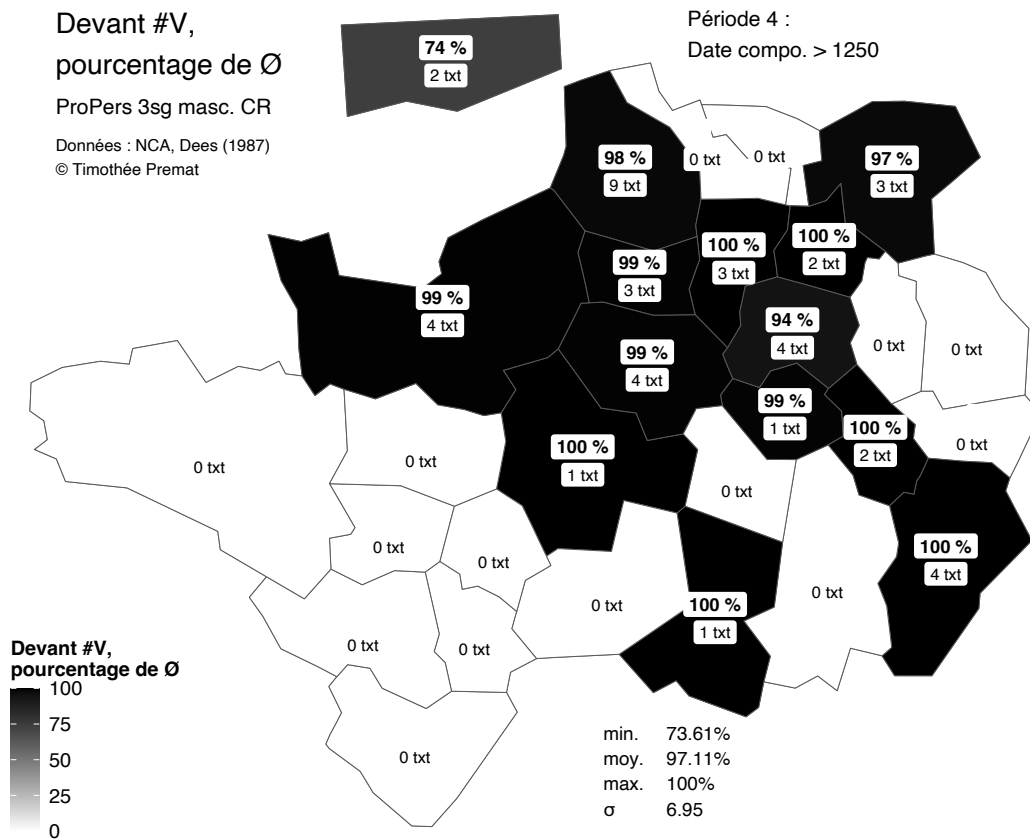
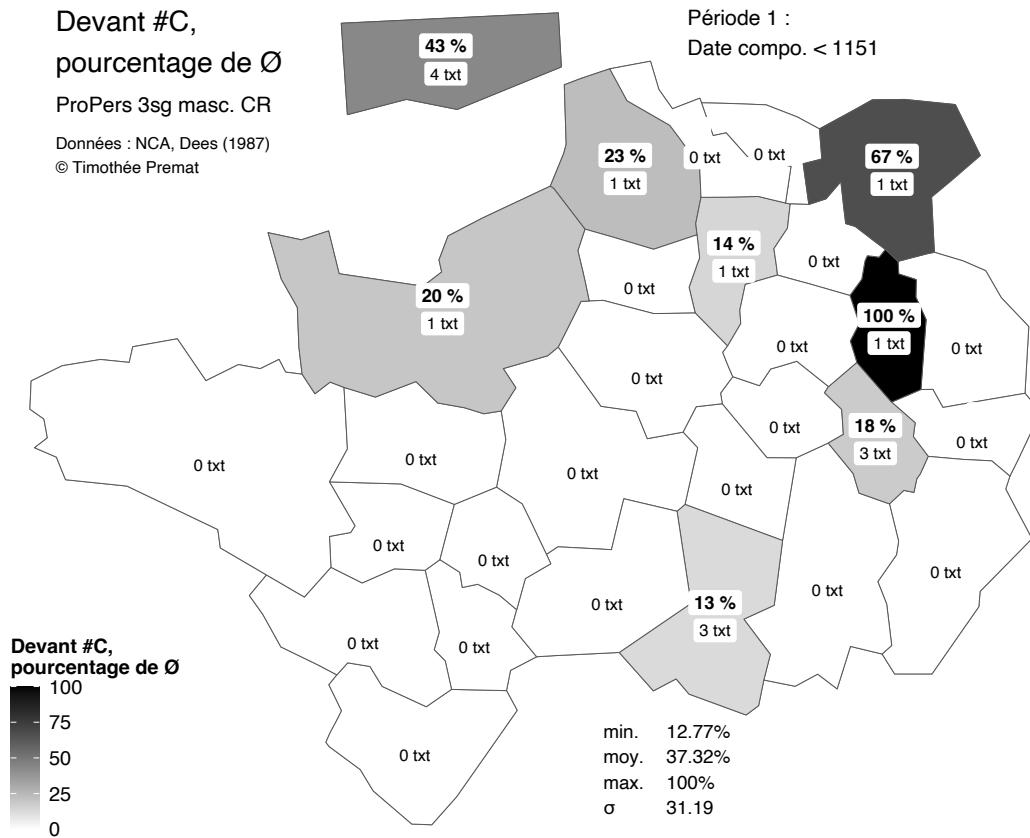


FIGURE E.9 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), pronoms personnels masculins régimes de troisième personne du singulier, devant initiale vocalique.

a



b

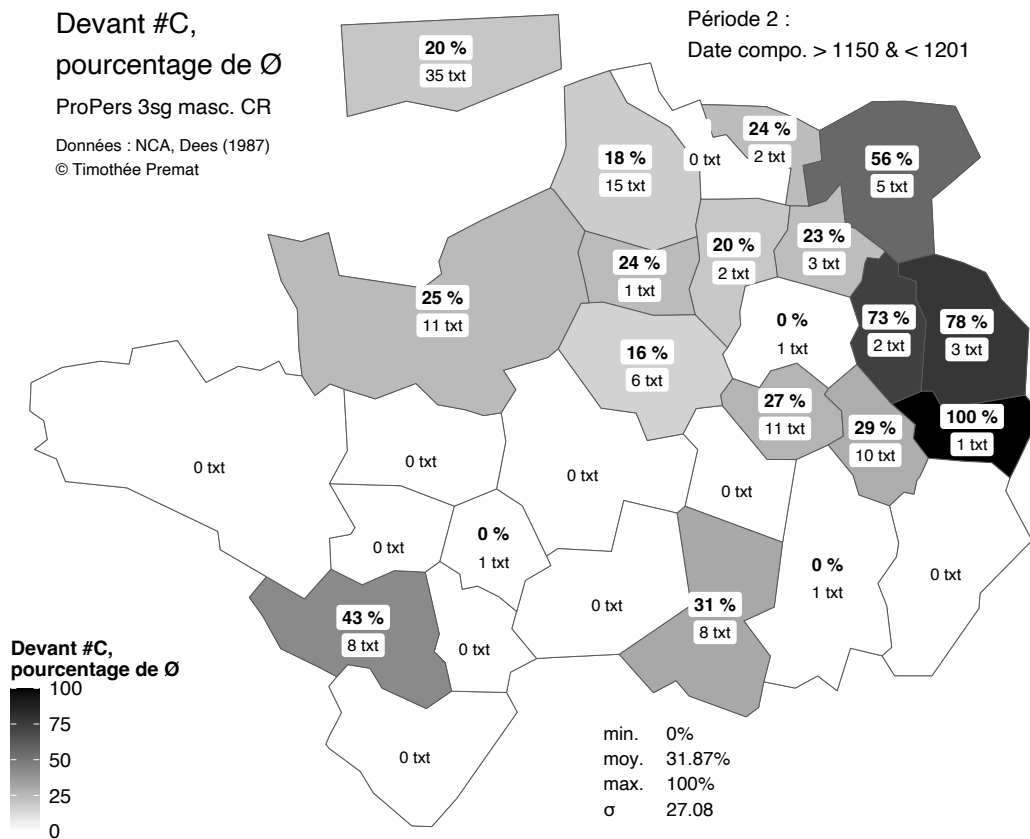
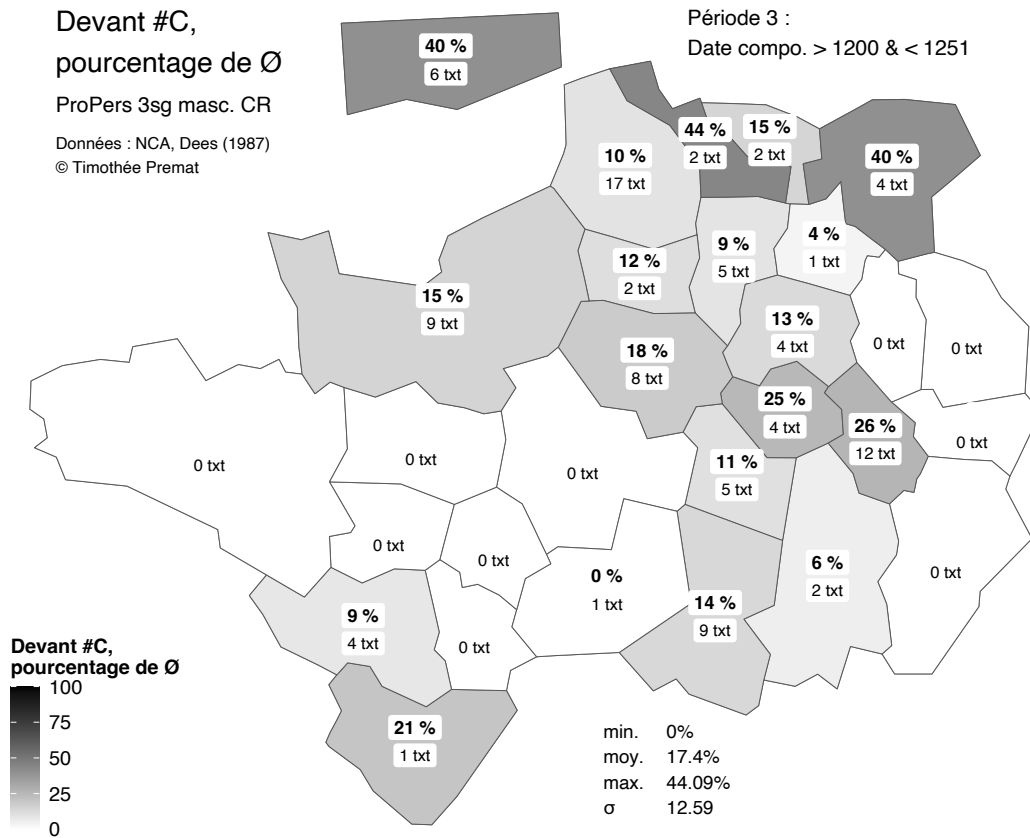


FIGURE E.10 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1050) et 2 (b : 1151-1200), pronoms personnels masculins régimes de troisième personne du singulier, devant initiale consonnantique.

a



b

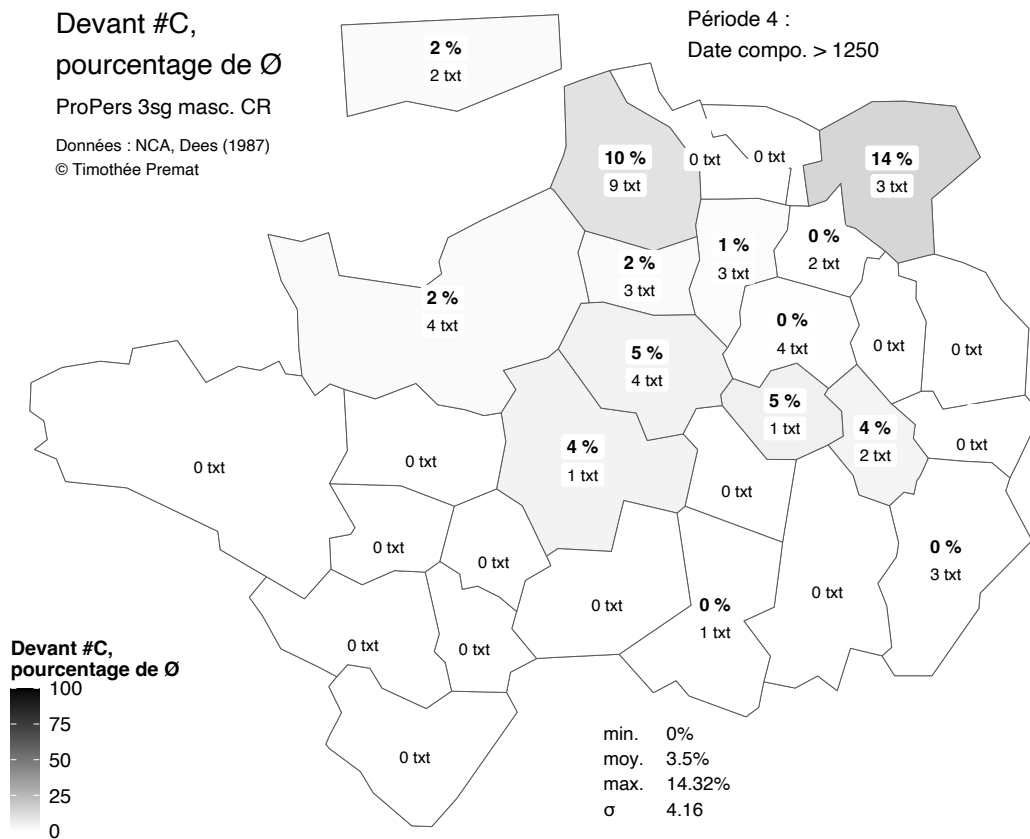
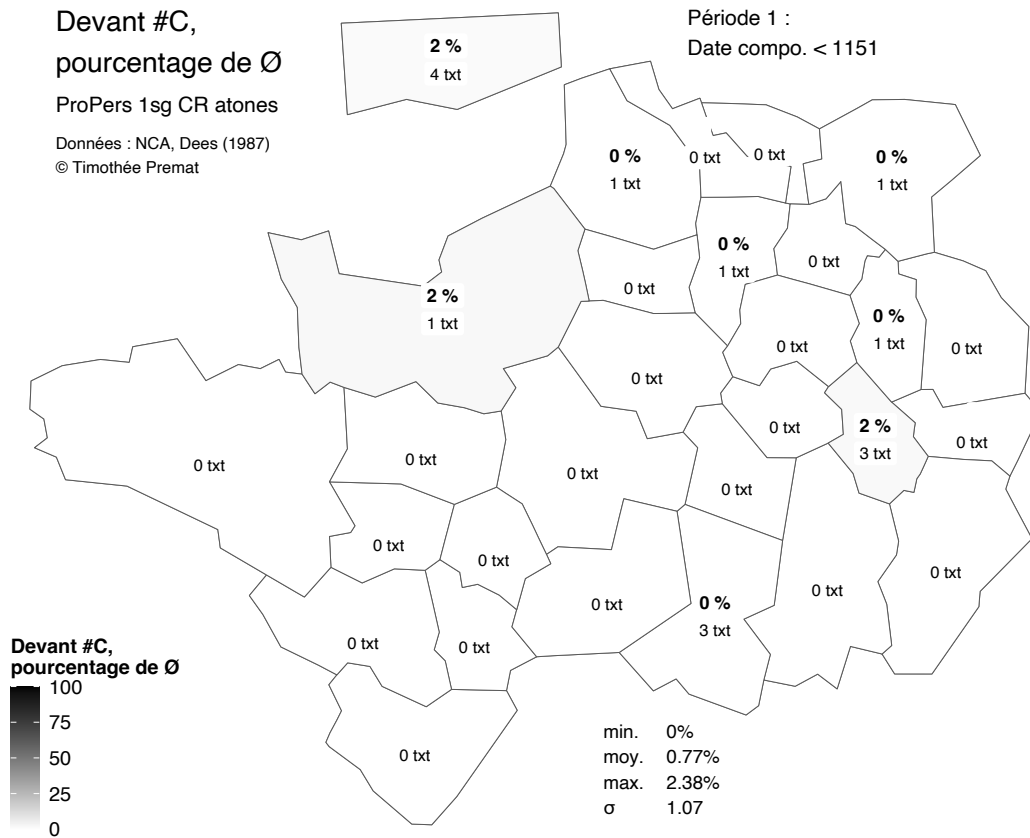


FIGURE E.11 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), pronoms personnels masculins régimes de troisième personne du singulier, devant initiale consonnante.

a



b

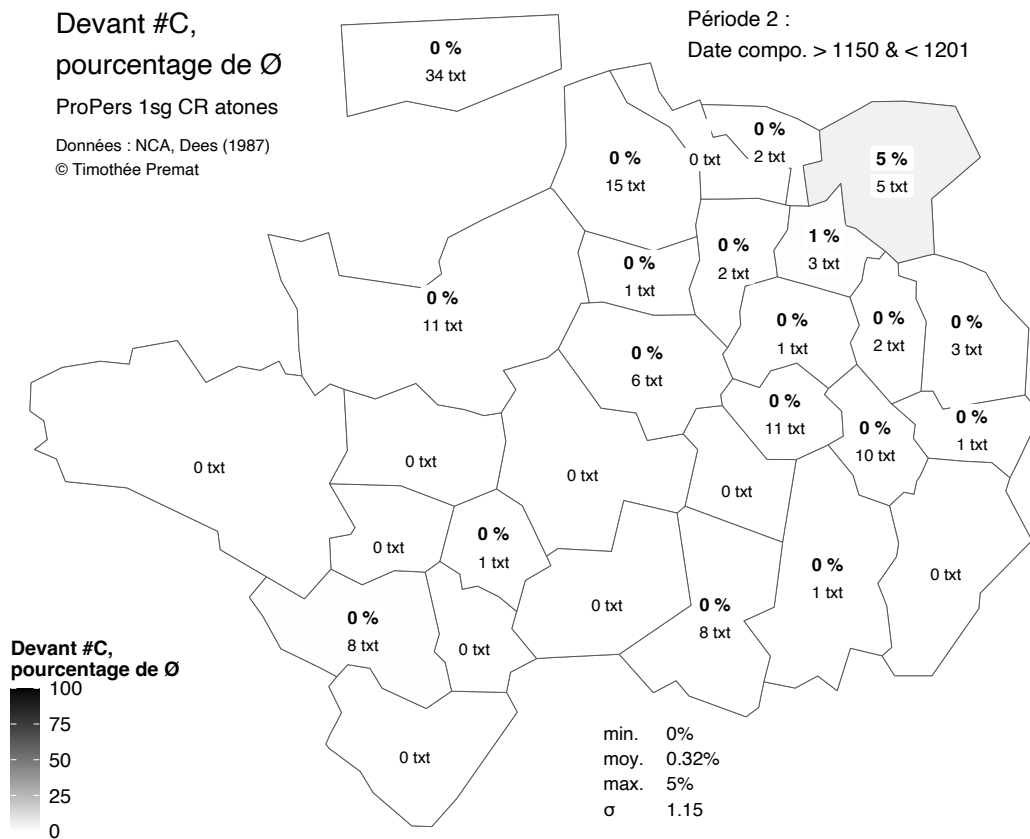
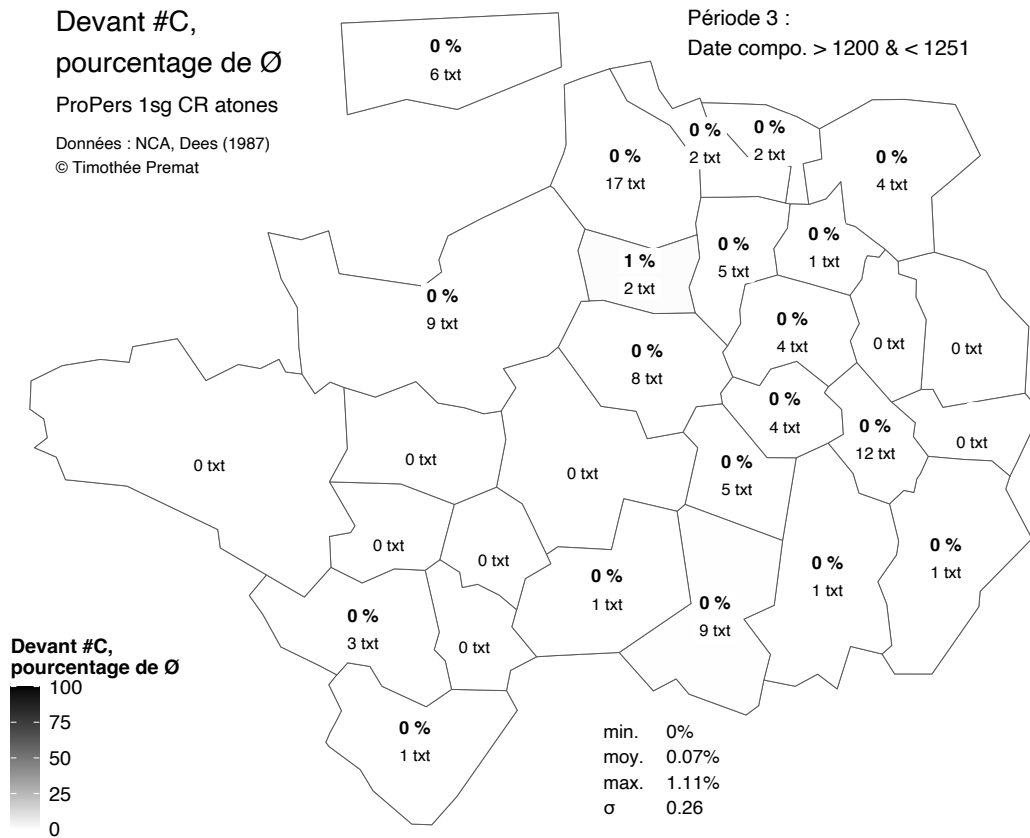


FIGURE E.12 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1050) et 2 (b : 1151-1200), pronoms personnels régimes de première personne du singulier, devant initiale consonantique.

a



b

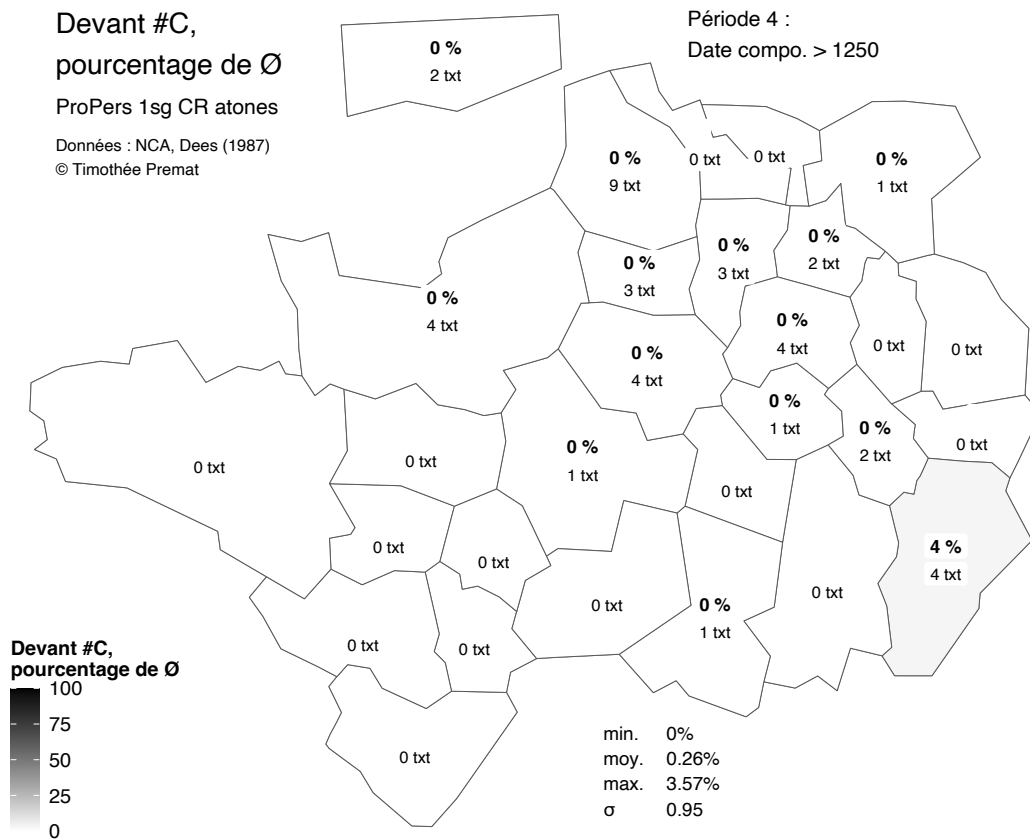
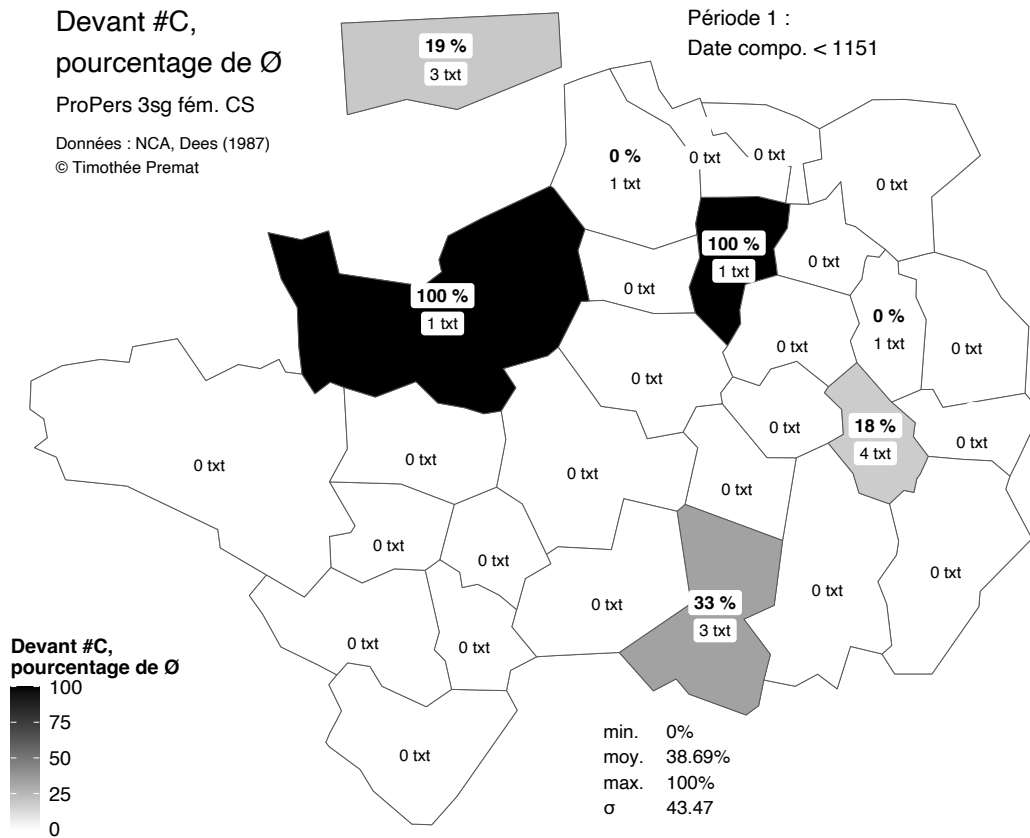


FIGURE E.13 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), pronoms personnels régimes de première personne du singulier, devant initiale consonnantique.

a



b

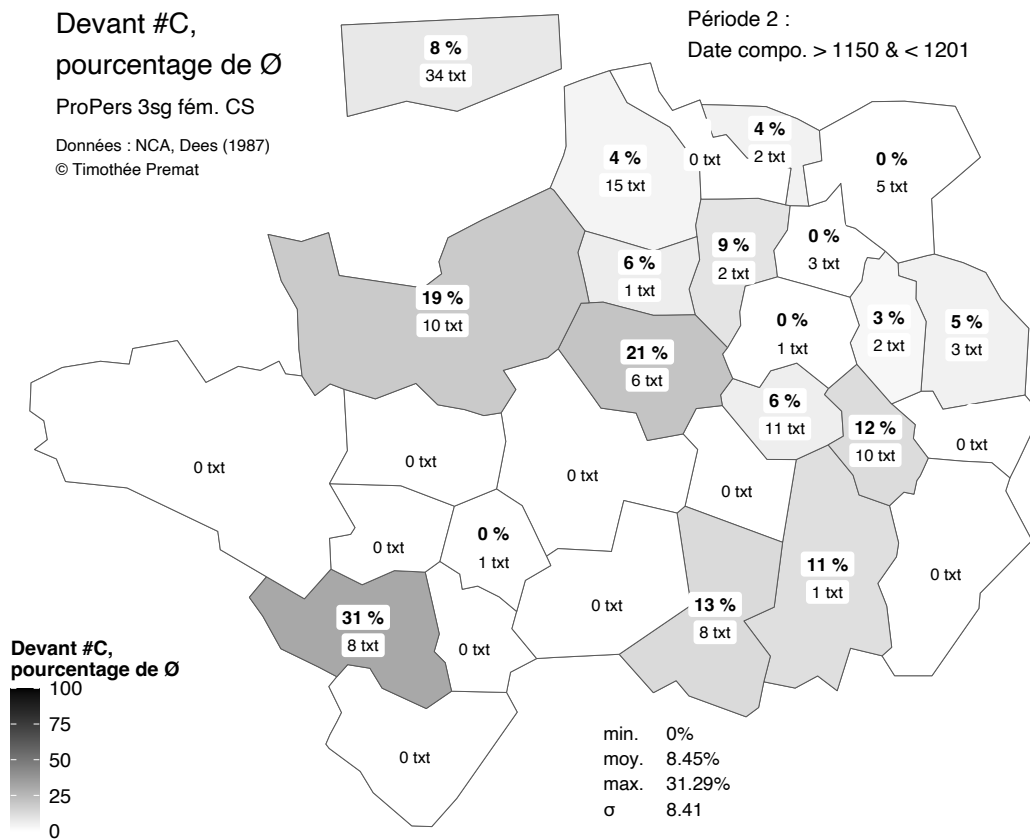
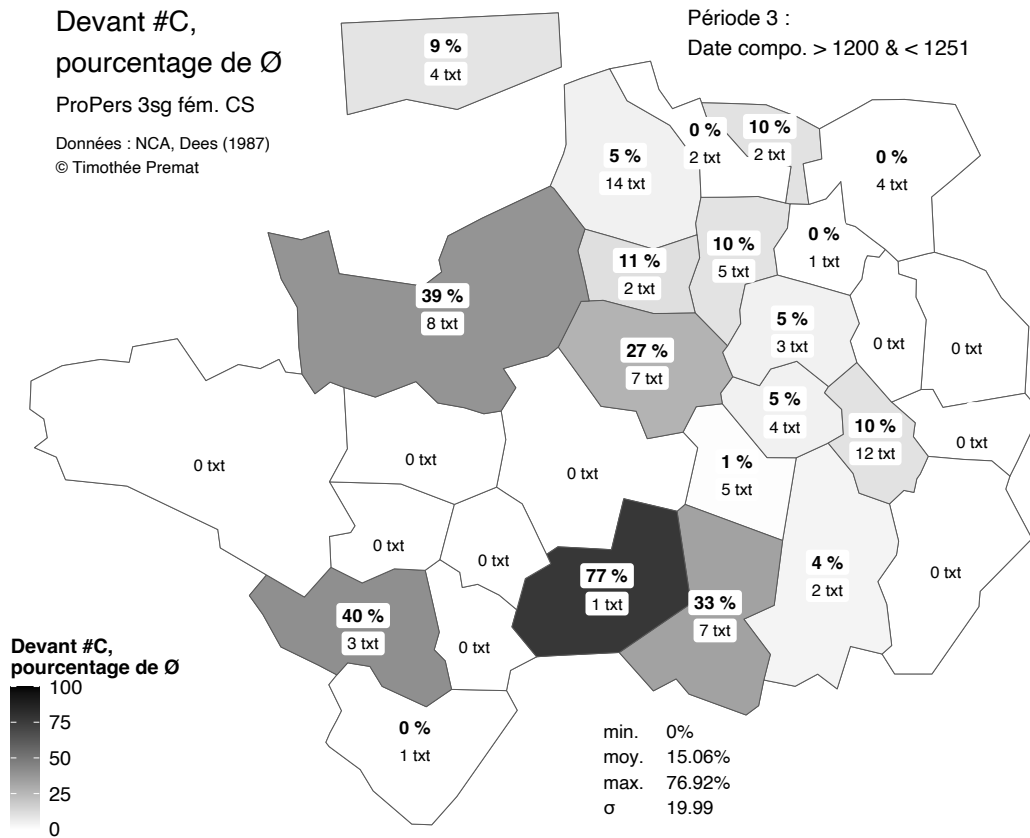


FIGURE E.14 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1050) et 2 (b : 1151-1200), pronoms personnels sujets féminins de troisième personne du singulier, devant initiale consonantique.

a



b

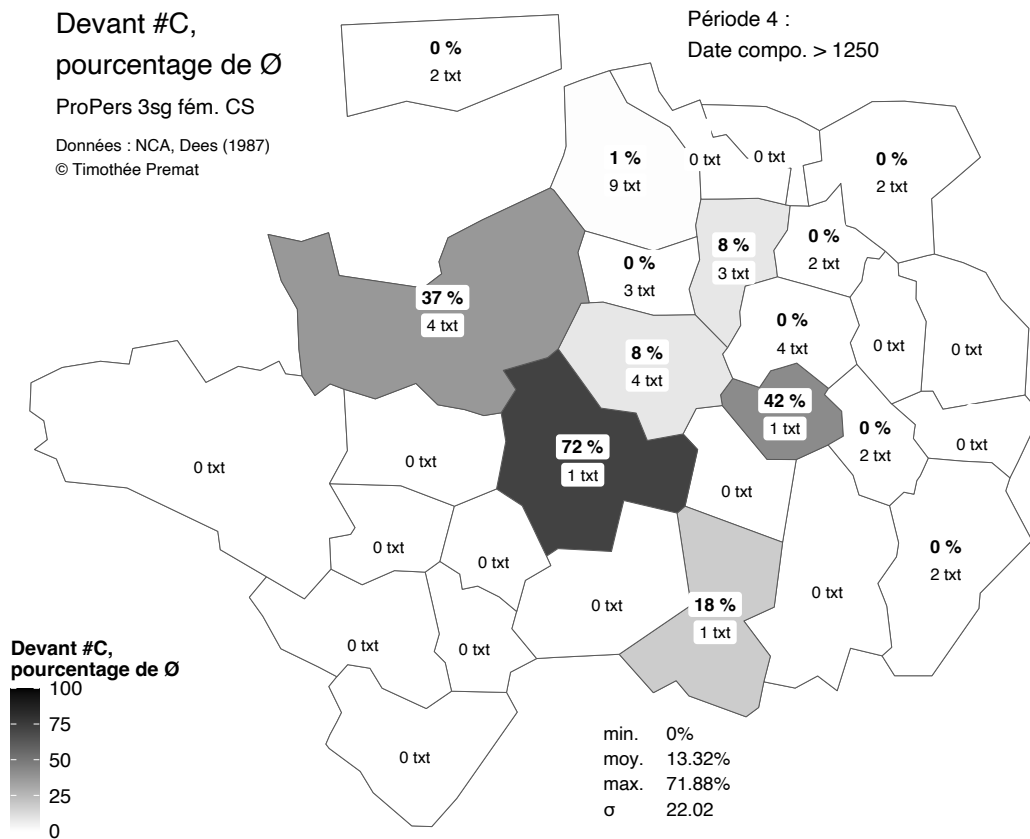


FIGURE E.15 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), pronoms personnels sujets féminins de troisième personne du singulier, devant initiale consonnante.

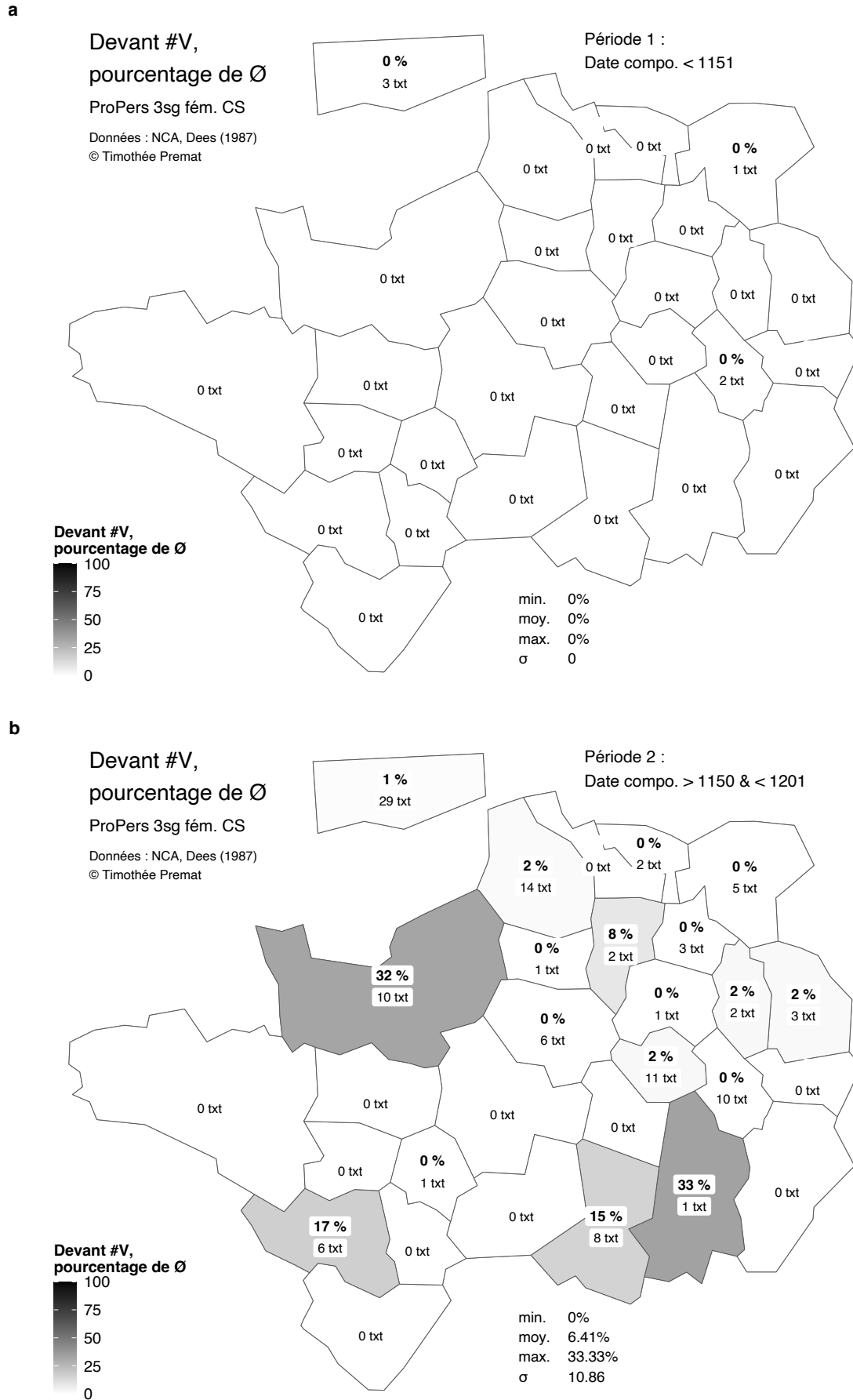
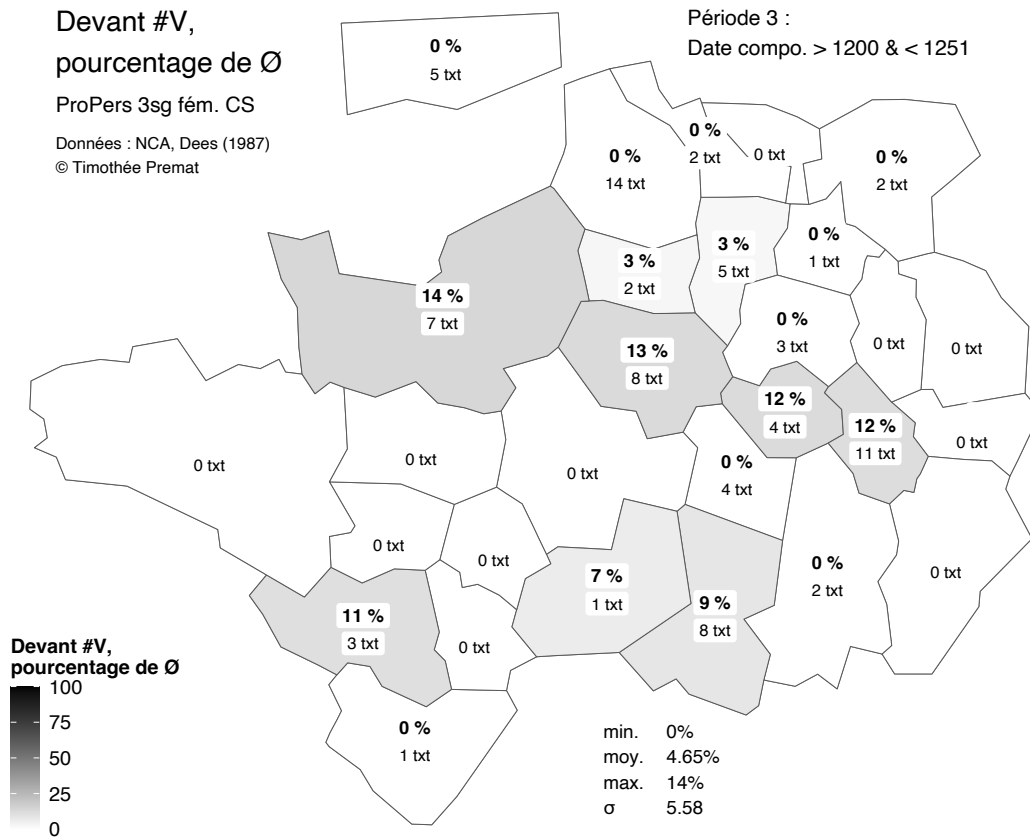


FIGURE E.16 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1050) et 2 (b : 1151-1200), pronoms personnels sujets féminins de troisième personne du singulier, devant initiale vocalique.

a



b

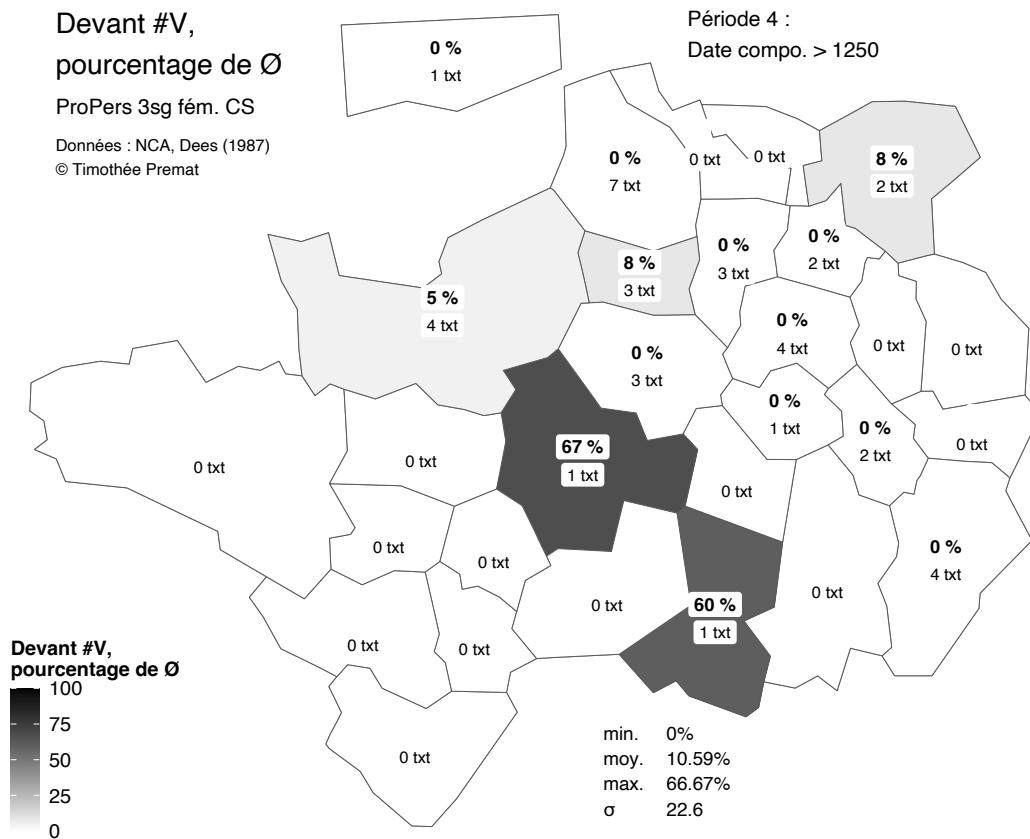
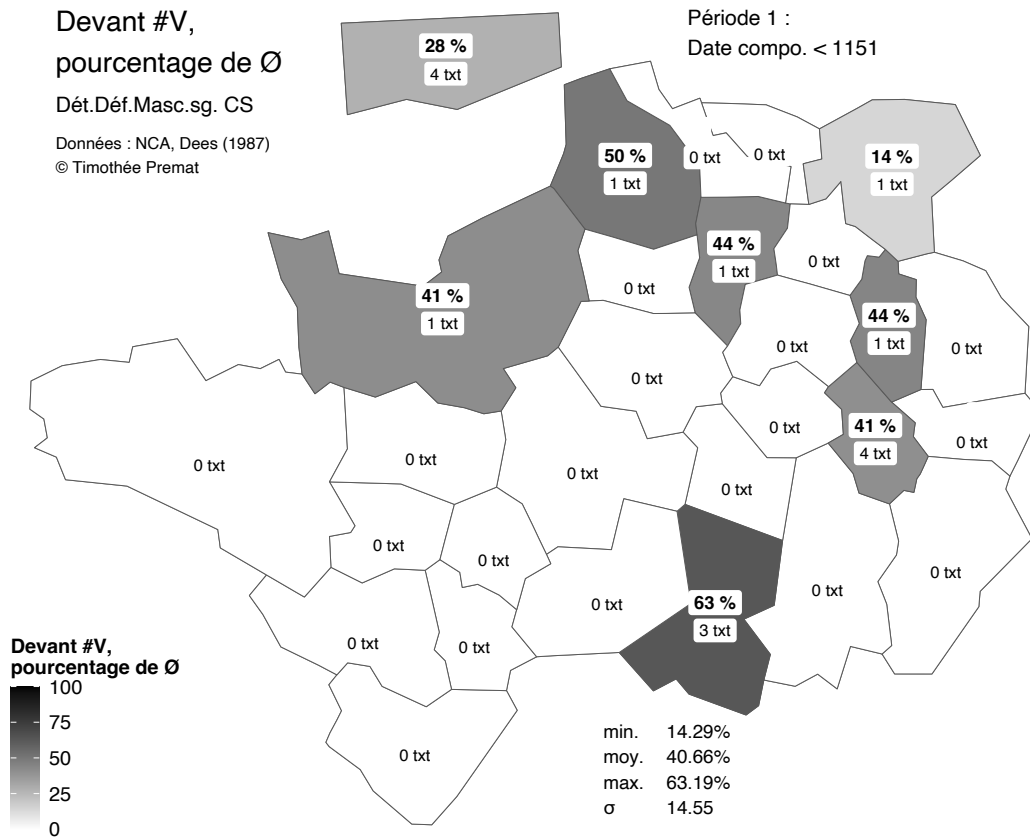


FIGURE E.17 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), pronoms personnels sujets féminins de troisième personne du singulier, devant initiale vocalique.

a



b

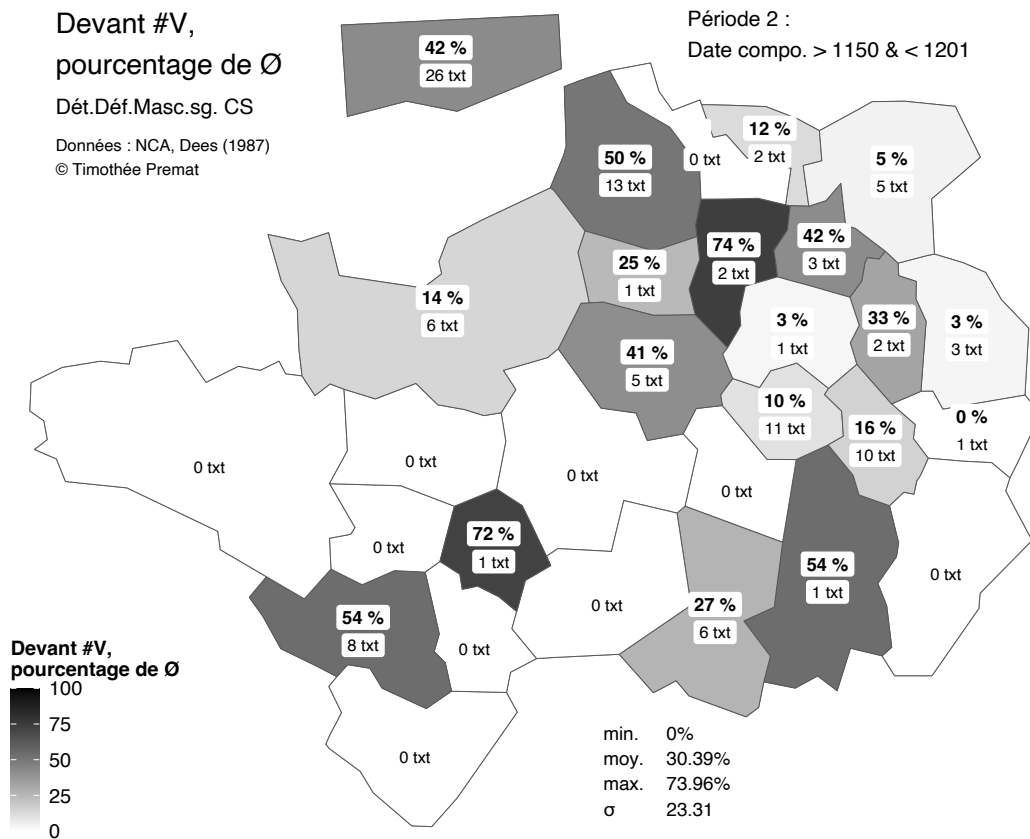
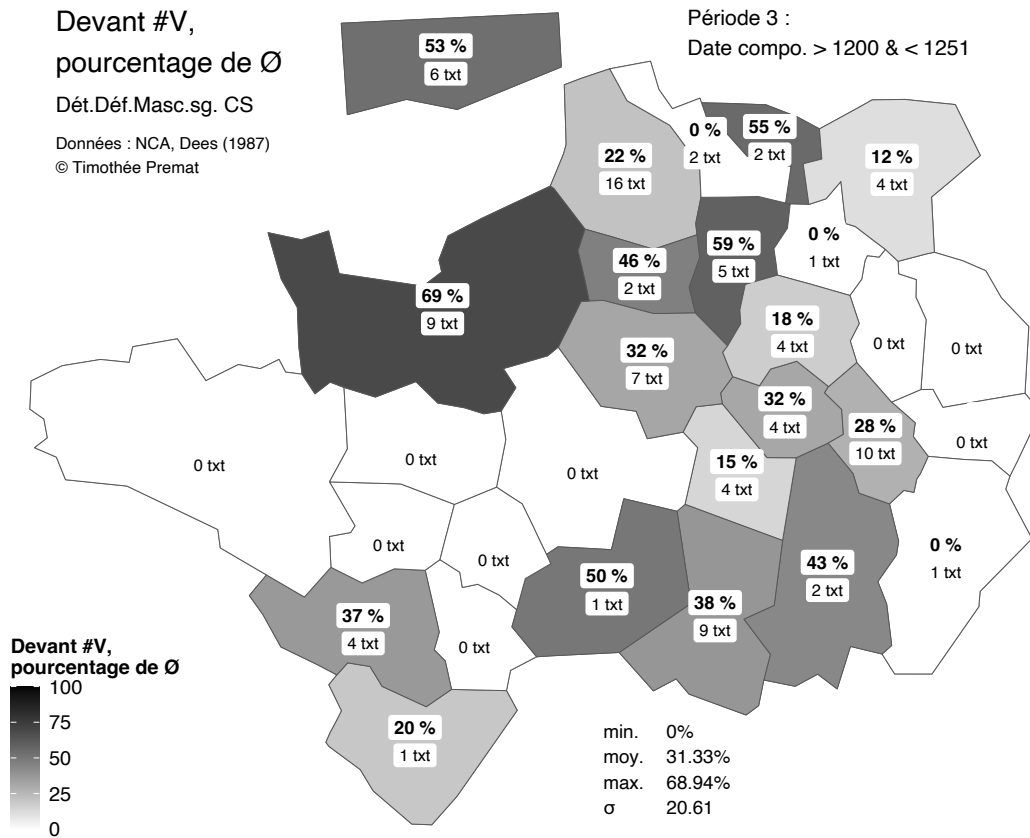


FIGURE E.18 – Distribution diatopique des alternances <V>/Ø, périodes 1 (a : 1100-1150) et 2 (b : 1151-1200), article défini masculin singulier au cas sujet (*li*), devant initiale vocalique.

a



b

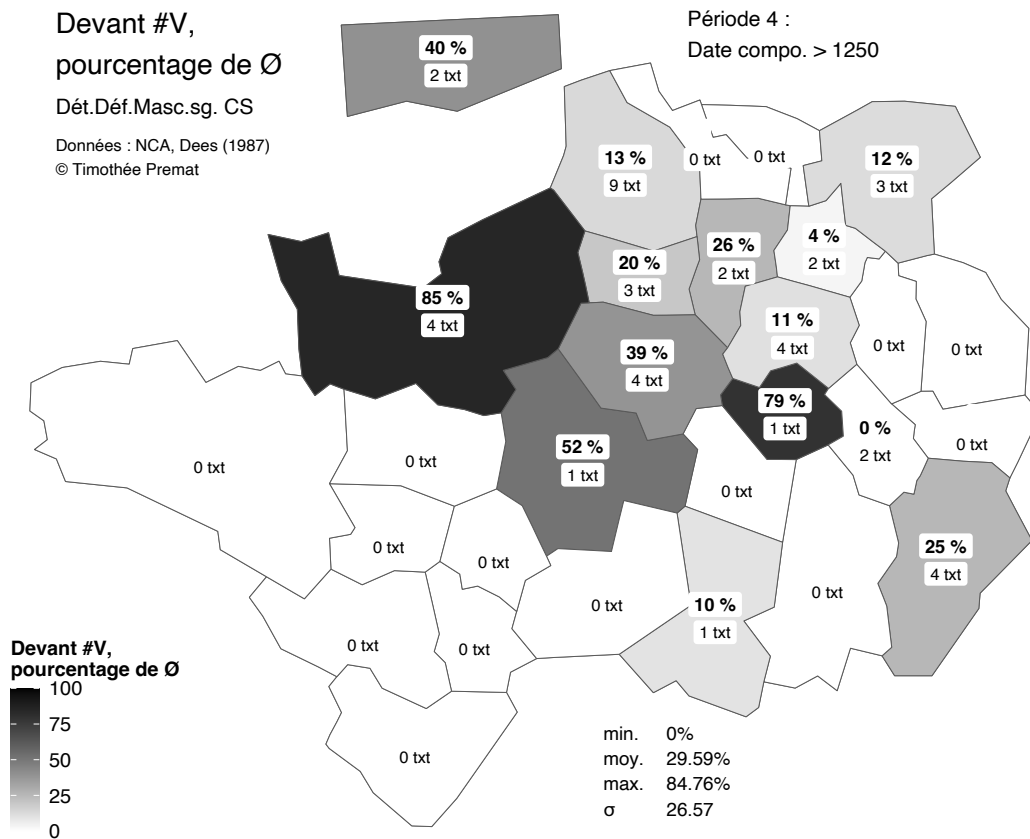
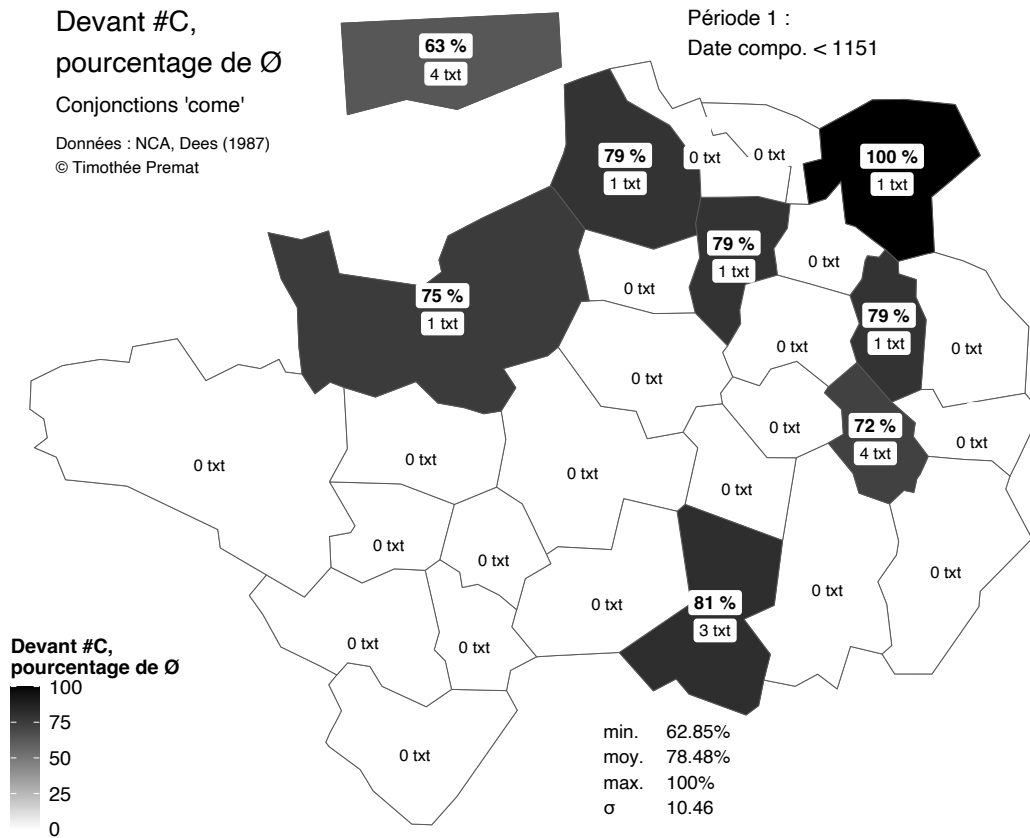


FIGURE E.19 – Distribution diatopique des alternances <V>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), article défini masculin singulier au cas sujet (*li*), devant initiale vocalique.

a



b

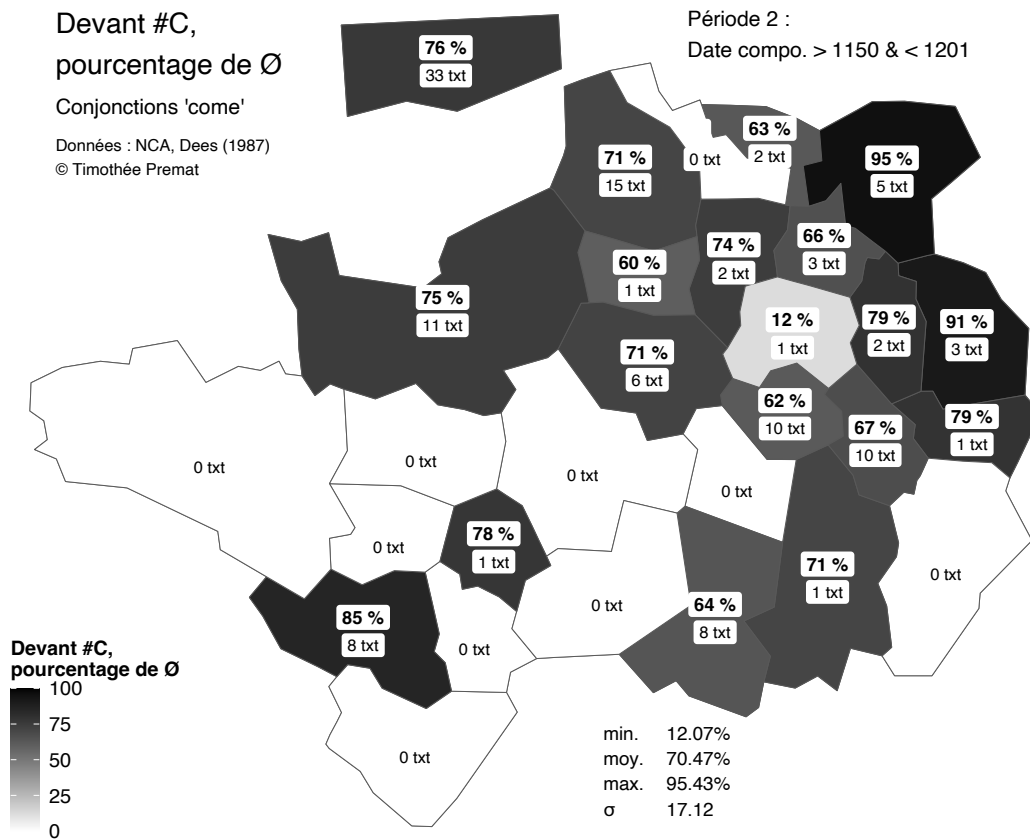
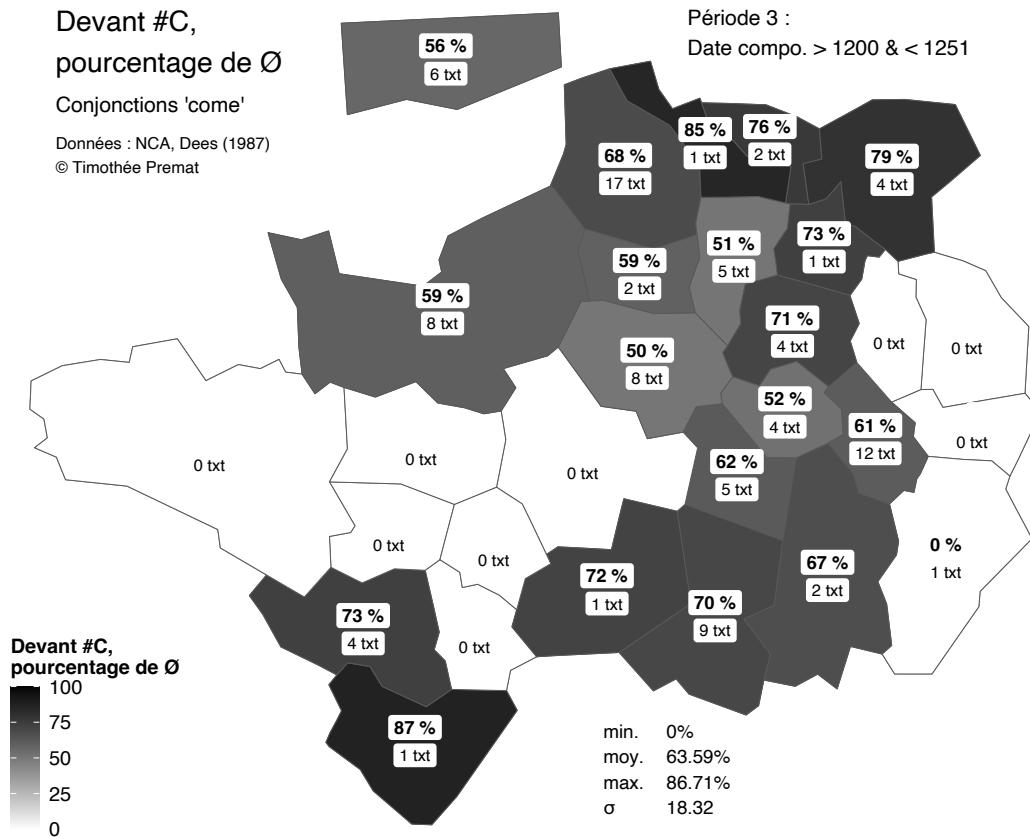


FIGURE E.20 – Distribution diatopique des alternances $\langle e \rangle / \emptyset$, périodes 1 (a : 1100-1050) et 2 (b : 1151-1200), conjonctions « come », devant initiale consonantique.

a



b

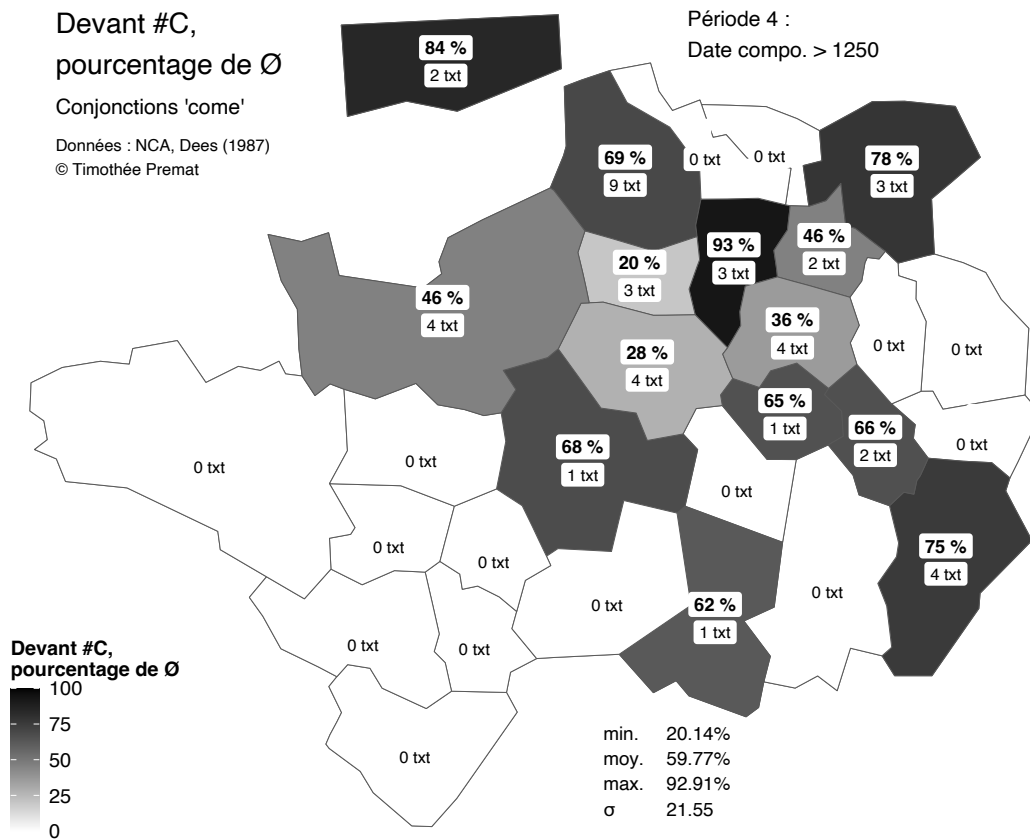
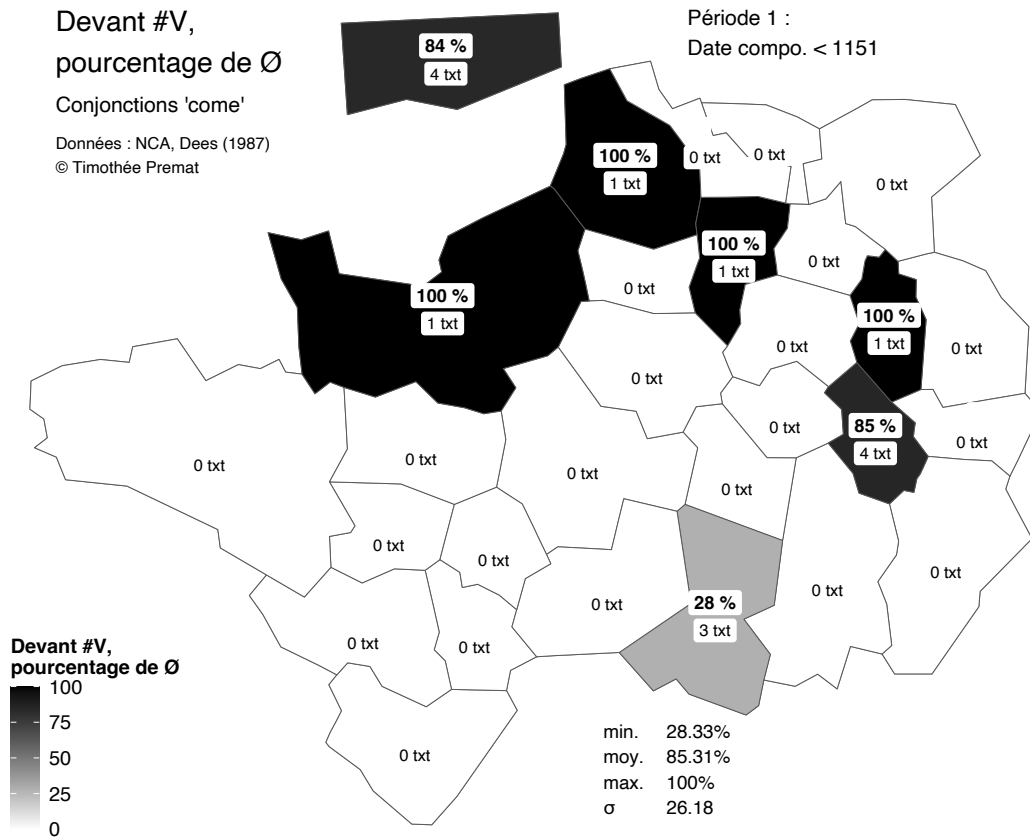


FIGURE E.21 – Distribution diatopique des alternances <e>/Ø, périodes 3 (a : 1201-1250) et 4 (b : 1251-1322), conjonctions « come », devant initiale consonantique.

a



b

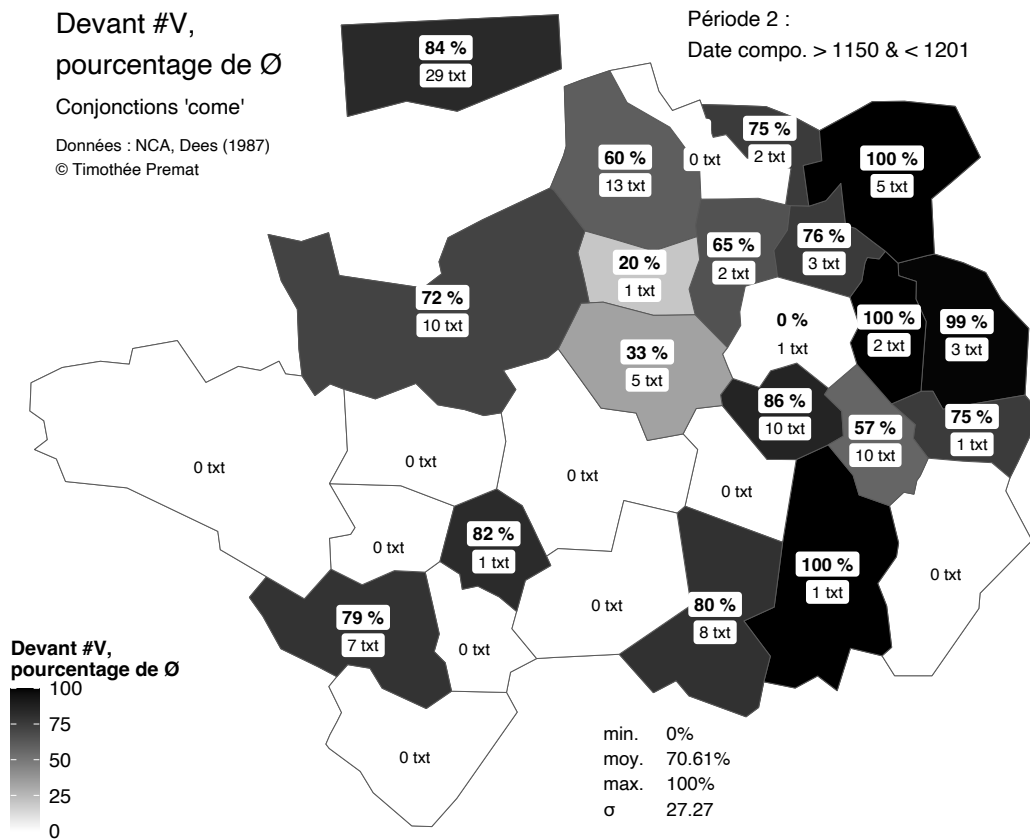


FIGURE E.22 – Distribution diatopique des alternances <e>/Ø, périodes 1 (a : 1100-1050) et 2 (b : 1151-1200), conjonctions « come », devant initiale vocalique.

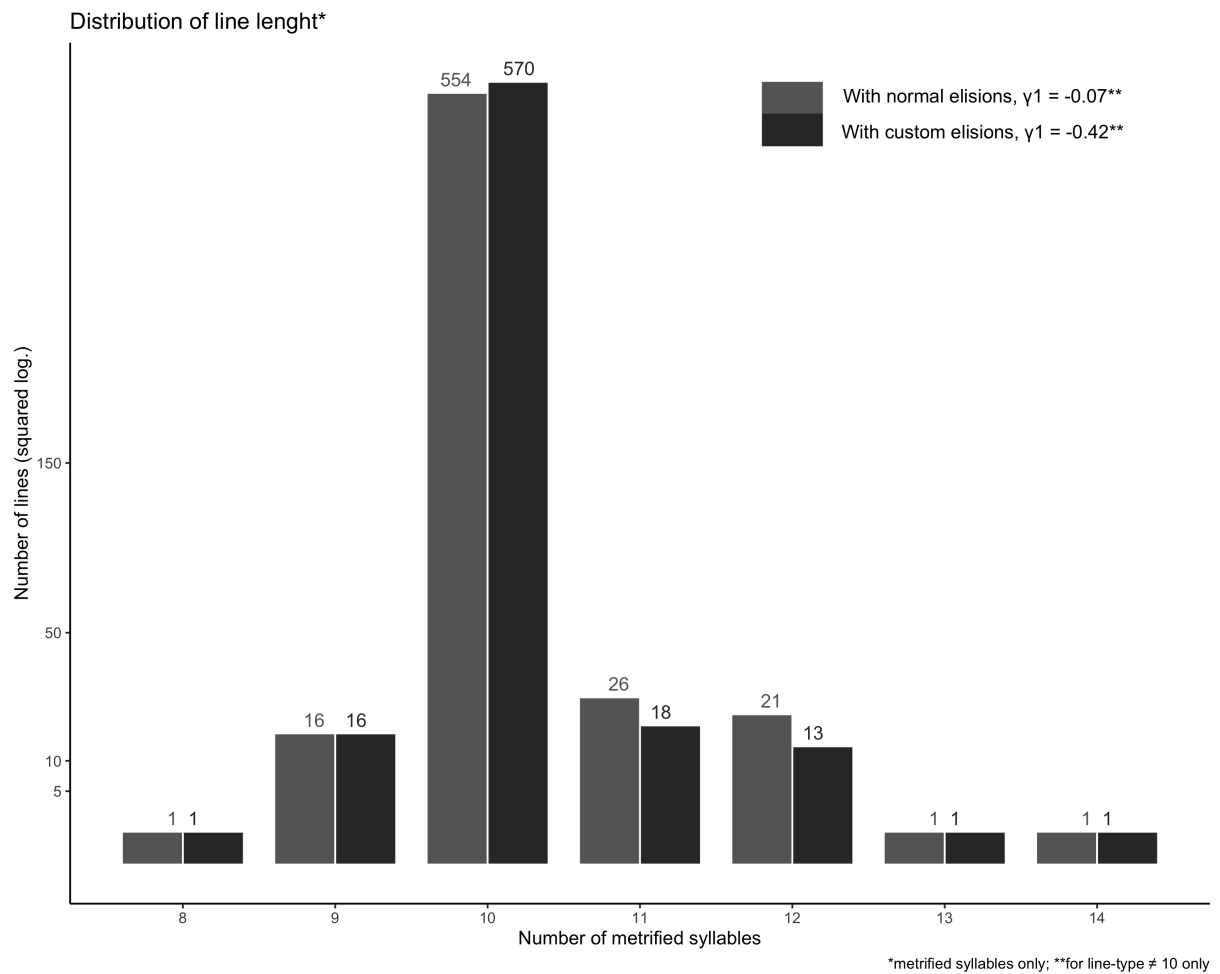


FIGURE E.24 – Distribution des longueurs de vers détectées par le PAM (config pour le t.a.fr.) dans *Alexis*. Le coefficient γ_1 mesure la symétrie du jeu de données, *normal elisions* (en gris) signifie que nous ne sommes pas intervenus sur l'élosion des monosyllabes, tandis que *custom elisions* (en noir) signifie que nous sommes intervenus sur les données pour corriger les cas d'élosions métriques non transcrites que nous considérons valides. La figure est issue de PREMATA et POGGIO (à paraître).

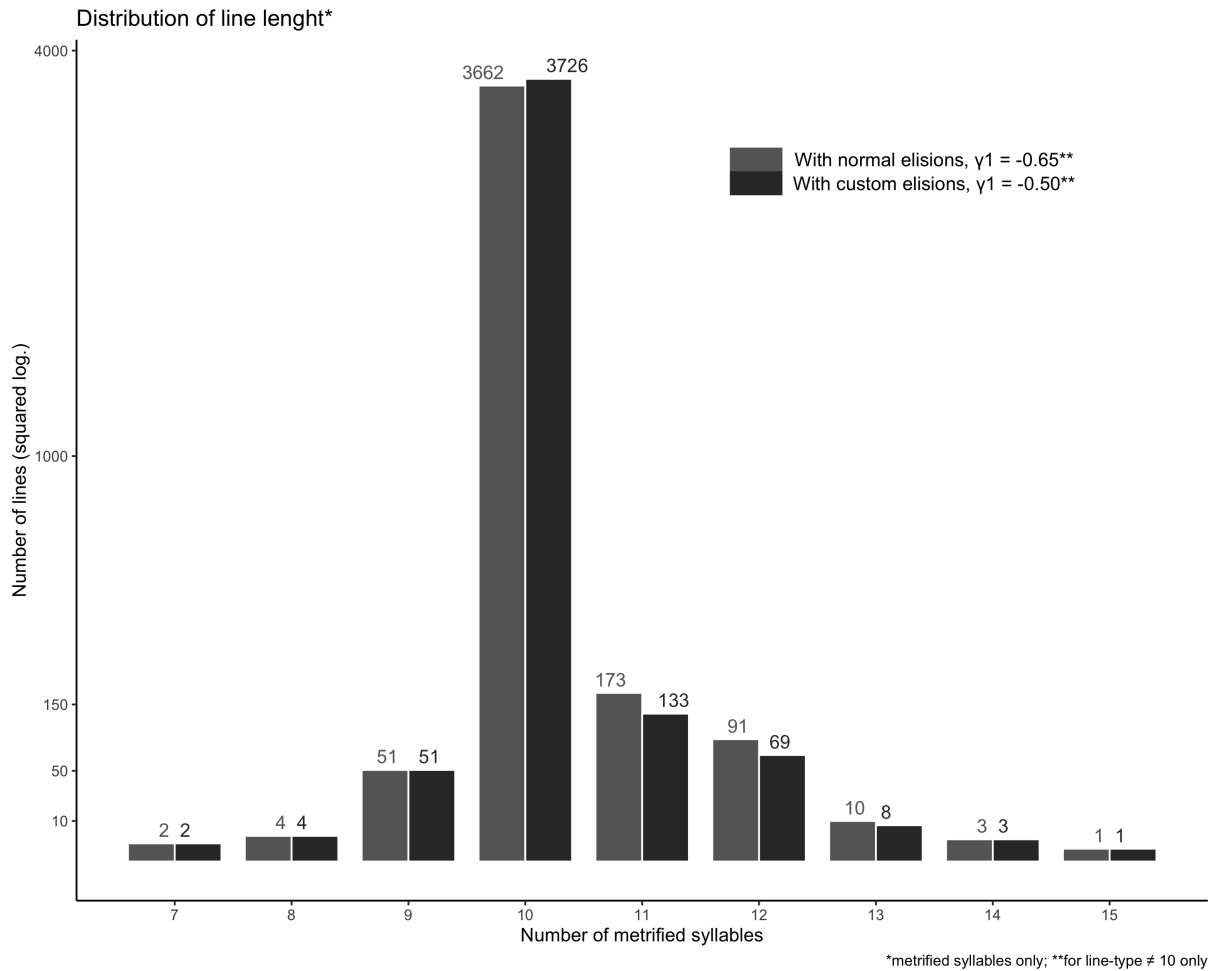


FIGURE E.25 – Distribution des longueurs de vers détectées par le PAM (conf ig pour le t.a.fr.) dans *Roland*. Le coefficient γ_1 mesure la symétrie du jeu de données, *normal elisions* (en gris) signifie que nous ne sommes pas intervenus sur l'élision des monosyllabes, tandis que *custom elisions* (en noir) signifie que nous sommes intervenus sur les données pour corriger les cas d'élisions métriques non transcrites que nous considérons valides. La figure est issue de PREMATA et POGGIO (à paraître).

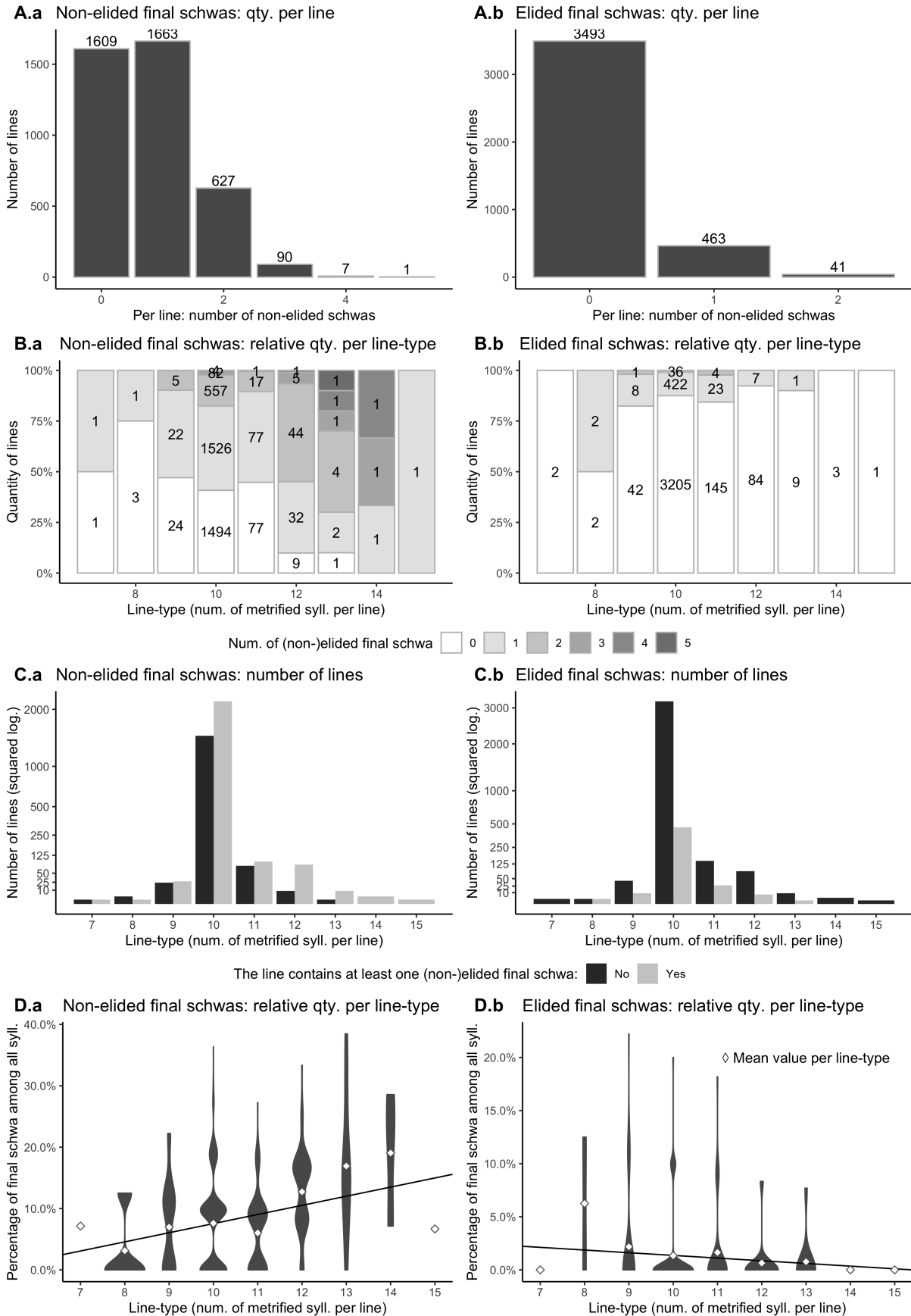


FIGURE E.26 – Ensemble de mesures exploratoires visant à isoler une potentielle corrélation entre la présence d’un ou plusieurs schwas finaux et l’hypo- ou hypermétrie des vers de Roland. L’augmentation du nombre de schwas par vers en fonction de la longueur des vers en D.a est conforme à une probabilité orthogonale que plus un vers est long, plus il peut comporter de schwas finaux.

Références

- ACADÉMIE FRANÇAISE (1718). *Nouveau Dictionnaire de l'Académie française dédié au Roy*. 2^e éd. 2 t. Paris : Jean-Baptiste Coignard.
- (1880 [1704]). « Observations de l'académie françoise sur les Remarques de M. de Vaugelas ». In : *Remarques sur la langue françoise*. 2 t. Versailles/Paris : Cerf et fils/J. Baudry [réédition des remarques de l'Académie, à l'origine sous forme autonome, ici sous forme d'appendice, par Alexis Chassang; url de l'original. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k50582j>.
- ADAM DE LA HALLE (1980). *Le gieus de Robin et de Marion*. Sous la dir. de René STUIP. Amsterdam : Festival de Hollande.
- (2002 [1998]). *Li Gieus de Robin et de Marion*. Sous la dir. d'Olivier BETTENS. URL : <https://virga.org/robin/>.
- ADAMS, Marianne (1987). *Old French, Null Subjects and Verb Second Phenomena*. Los Angeles : PhD thesis, University of California.
- ANDOLF, Sven, éd. (1965). *Floovant, Chanson de Geste du XIII^e siècle*. Uppsala : Almqvist & Wiksells.
- Anglo-Norman Dictionary (AND² Online Edition)* (2022). Aberystwyth University. URL : <https://anglo-norman.net/>.
- ANON. (1882). *Lyoner Yzopet : altfranzösische Übersetzung des XIII. Jahrhunderts in der Mundart der Franche-Comté, mit dem kritischen Text des lateinischen Originals*. Sous la dir. de Wendelin FOERSTER. Hellbronn : Gebr. Henninger.
- (1899). *Orson de Beauvais, chanson de geste du XII^e s.* Sous la dir. de Gaston PARIS. Paris : Firmin Didot.
- (1903). *Robert le Diable, roman d'aventure*. Sous la dir. d'Eilert LÖSETH. Paris : F. Didot.
- (1937). *The medieval French Roman d'Alexandre*. Sous la dir. de Milan Sylvanus LA DU, Edward Cooke ARMSTRONG et Alfred FOULET. Princeton/Paris : Princeton University Press/Presses Universitaires de France.
- (1955). *Le chevalier au barisel : conte pieux du XIII^e siècle*. Sous la dir. de Félix LECOY. Paris : Honoré Champion.
- (1963). *Le mystère d'Adam — ordo representacionis ade*. Sous la dir. de Paul AEBISCHER. Genève/Paris : Droz/Minard.
- (1967). *Aye d'Avignon : chanson de geste anonyme*. Sous la dir. de S. J. BORG. Genève : Droz.
- (1974). *Amadas et Ydoine : roman du XIII^e siècle*. Sous la dir. de John Revell REINHARD. Paris : Honoré Champion.
- (1969 [1921]). *Le garçon et l'aveugle : jeu du XIII^e siècle*. Sous la dir. de Mario ROQUES. Paris : Honoré Champion.
- ANONYME (ca. 1525 [1902]). « L'Art et Science de Rhétorique ». In : *Recueil d'arts de seconde rhétorique*. Sous la dir. d'Ernest LANGLOIS. Paris : Imprimerie nationale, p. 265-426.
- AROUI, Jean-Louis (2019). « Césure et parophonologie en français, en grec et en latin ». Présentation au *Séminaire de métrique générale*, Paris, Sorbonne Université, 25/04/2019.
- AURACHER, Theodor M., éd. (1877). *Die sogenannte poitevinische Uebersetzung des Pseudo-Turpin*. Halle : M. Niemeyer.
- AVALLE, D'Arco Silvio, éd. (1962). *Passion de Jésus-Christ ou Passion de Clermont*. Lyon : ENS de Lyon. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 03-12-2016. URL : <http://catalog.bfm-corporus.org/passion>.
- éd. (1967). *Vie de saint Léger*. Pavia : Tipografia del libro. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 3-12-2016. URL : <http://catalog.bfm-corporus.org/slethgier>.
- BADIOU-MONFERRAN, Claire (2020). « Les coordonnants *et, ou, ni* et les locutions coordonantes ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 944-961.
- BAILLY, Anatole et Émile EGGER (1935). *Dictionnaire grec-français*. 11^e éd. Paris : Hachette. URL : <https://archive.org/embed/BaillyDictionnaireGrecFrancais>. Cf. GRÉCO, CHARBONNET, DE WILDE et MARÉCHA (2020).
- BAKER, Alfred T. et Alexander BELL, éd. (1947). *St. Modwenna*. London/Oxford : Anglo-Norman Texts Society.
- BAKER, J-A.T (1929). « La Vie de saint Edmond, archevêque de Cantorbéry ». In : *Romania* 55.219-220, p. 332-381.
- BALLY, Charles (1944). *Linguistique Générale et Linguistique Française*. 2nd édition entièrement refondue ; première édition 1932, Paris : PUF/Ernest Leroux. Berne : Francke.

- BAYOT, Alphonse, éd. (1929). *Le Poème Moral, traité de la vie chrétienne, écrit dans la région wallonne vers l'an 1200*. Bruxelles/Liège : Académie Royale de Langue et de Littérature françaises de Belgique.
- éd. (1931). *Gormont et Isembart, Fragment de chanson de geste du XII^e s.* 2nd édition revue; première édition 1914. Paris : Honoré Champion.
- BEAULIU, Guischart de (1909). *Le Sermon de Guischart de Beauliu*. Sous la dir. d'Arvid GABRIELSON. Uppsala/-Leipzig : Akademiska Bokhandeln/Harrassowitz.
- BEAVER, Joseph C. (1968). « Progress and Problems in Generative Metrics ». In : *Papers from the Fourth Regional Meeting (Chicago Linguistic Society)*. Sous la dir. de Bill J. DARDEN, Charles James Nice BAILEY et Alice DAVISON. Chicago : Chicago Linguistic Society, p. 146-155.
- BÉDIER, Joseph, éd. (1947). *La chanson de Roland, publiée d'après le manuscrit d'Oxford et traduite par Joseph Bédier*. réédition posthume, les éditions précédentes ne sont pas qualifiées de *définitives* dans le titre ou sur le colophon; première édition : 1922. Paris : Édition d'art H. Piazza.
- BENDJABALLAH, Sabrina et Martin HAIDEN (2008). « A Typology of Emptiness in Templates ». In : *Sounds of Silence: Empty Elements in Syntax and Phonology*. Sous la dir. de Jutta HARTMANN, Veronika HEGEDÜS et Hank van RIEMSDIJK. Amsterdam : Elsevier, p. 23-59.
- BENOÎT DE SAINTE-MAURE (1904). *Le roman de Troie*. Sous la dir. de Léopold CONSTANS. Paris : F. Didot.
- BENOIT DE SAINTE-MAURE (1951-1967). *Chronique des ducs de Normandie*. Sous la dir. de Carin FAHLIN. T. 1. 4 t. Uppsala : Almqvist & Wiksell.
- BERGOUNIOUX, Gabriel (1989). « Le francien (1815-1914) : la linguistique au service de la patrie ». In : *Mots* 19, p. 23-40.
- BERKEY, Max L. (1965). « *The Liber sancti Jacobi* : The French Adaptation by Pierre de Beauvais ». In : *Romania* 341.86, p. 77-103.
- BERNARD DE CLAIRVAUX (1894). *Predigten des h. Bernhard in altfranzösischer Übertragung : aus einer Handschrift der königlichen Bibliothek zu Berlin*. Sous la dir. d'Alfred SCHULZE. Tübingen : Litterarischen Verein.
- BIRD, Steven, Ewan KLEIN et Edward LOPER (2009). *Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit*. Sebastopol/Beijing/Cambridge : O'reilly.
- BLAKEY, Brian (1968). « *Aucassin et Nicolette*, xxiv, 4 ». In : *French Studies* 22, p. 97-98.
- BLANCHE-BENVENISTE, Claire, Denis APOTHÉLOZ, Marie-José BÉGUELIN, Christophe BENZITOUN, Gilles CORMINBOEUF, José DEULOFEU, Peter LAUWERS, Dominique WILLEMS, Alain BERRENDONNER et Jeanne-Marie DEBAISIEUX (2020 [2018]). « La paradigmatique : statut des catégories ». In : *Encyclopédie grammaticale du français*. Sous la dir. de Claire BLANCHE-BENVENISTE, Denis APOTHÉLOZ, Marie-José BÉGUELIN, Christophe BENZITOUN, Gilles CORMINBOEUF, José DEULOFEU, Peter LAUWERS, Dominique WILLEMS, Alain BERRENDONNER et Jeanne-Marie DEBAISIEUX. [en ligne], version du 04/2020. URL : <http://www.encyclogram.fr>.
- BLUMENFELD, Lev (2016). « Generative Metrics : An Overview ». In : *Language and Linguistics Compass* 10.9, p. 413-430. DOI : <https://compass.onlinelibrary.wiley.com/doi/10.1111/lnc3.12204>.
- BONNARDOT, François (1876). « Dialogus anime conquerentis et rationis consolantis, traduction en dialecte lorrain du XII^e s. ». In : *Romania* 5, p. 269-332. URL : <https://www.jstor.org/stable/45042735>.
- BOOIJ, Geert (1986). « Cliticization as prosodic integration: The case of Dutch ». In : *The Linguistic Review* 13, p. 219-242.
- BORETIUS, Alfred et Victor KRAVSE, éd. (1887). *Capitularia Regum Francorum*. T. 2. coll. *Monumenta Germaniae historica*. Hanovre : Impensis bibliopolii Hahniani.
- BOUCHERIE, Anatole (1873). *Le dialecte poitevin au XIII^e siècle*. Paris : Pedone-Lauriel.
- BOURCIEZ, Édouard (1955). *Précis historique de phonétique française*. 8^e édition, revue et corrigée. Paris : Klincksieck.
- (1967 [1910]). *Éléments de linguistique romane*. 5^e édition. Paris : Klincksieck.
- BOURCIEZ, Édouard et Jean BOURCIEZ (1974 [1967]). *Phonétique française, Étude historique*. 2^e éd. Paris : Klincksieck.
- BOZON, Nicholas (1905). *La Plainte d'amour, poème anglo-normand*. Sous la dir. de Johan VISING. Göteborg : Zachrisson.
- BRITAIN, Margaret S. (1900). *Historical Primer of French Phonetics and Inflection*. Oxford : Clarendon.
- BRUNEL, Clovis, éd. (1926). *La fille du comte de Pontieu, nouvelle du XIII^e siècle*. Paris : Champion.
- BYBEE, Joan (2000). « The phonology of the lexicon: Evidence from lexical diffusion ». In : *Usage-based models of language*. Sous la dir. de Michael BARLOW et Suzanne KEMMER. Stanford : CSLI Publications, p. 65-85.

- BYBEE, Joan (2002). « Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change ». In : *Language Variation and Change* 14, p. 261-290.
- (2006). « From Usage to Grammar: The Mind's Response to Repetition ». In : *Language* 82.4, p. 711-733.
- BYBEE, Joan et Joanne SCHEIBMAN (1999). « The effect of usage on degrees of constituency: the Reduction of *don't* in English ». In : *Linguistics* 37.4, p. 575-596.
- CARLIER, Anne (2020). « Les articles définis, indéfini et partitifs ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 664-674.
- CARLIER, Anne et Michèle GOYENS (1998). « De l'ancien français au français moderne : régression du degré zéro de la détermination et restructuration du système des articles ». In : *Cahiers de l'Institut de Linguistique de Louvain-la-Neuve* 24.3-4, p. 77-112.
- CARLIER, Anne, Céline GUILLOT-BARBANCE, Christiane MARCHELLO-NIZIA et Lene SCHØSLER (2020). « Catégories variables : noms, adjectifs, pronoms et déterminants ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 632-745.
- CAZAL, Yvonne, Gabriella PARUSSA et Elena LLAMAS-POMBO (2020). « Graphies : des usages à la norme ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 501-549.
- CHAMBERS, J. K. et Peter TRUGDILL (1998). *Dialectology*. Cambridge/New York/Melbourne : Cambridge University Press.
- CHANTRAINE, Pierre (1968). *Dictionnaire étymologique de la langue grecque. Histoire des mots*. 1^{re} éd. 2 t. Paris : Klincksieck.
- CHAO, Yuen-Ren (1934). « The non-uniqueness of phonemic solutions of phonetic systems ». In : *Bulletin of the Institute of History and Philology* 4.4, p. 219-224.
- CHASLE, Nathalie (2008). « Manifestation de la latence en ancien français aux x^e et xi^e siècles : liaison et redoublement syntaxique ». In : *Congrès Mondial de Linguistique Française - CMLF 2008*, p. 1645-1656. DOI : [10.1051/cmlf08175](https://doi.org/10.1051/cmlf08175).
- CHOUVION, Sophie, Timothée PREMAT et Axelle VERNER (2021). « La musication dans les chansons d'Henry Fresneau, Grammaire d'association, paraphonologie et forme polyphonique au xvi^e siècle ». In : *Textus & Musica [en ligne]* 3. URL : <https://textus-et-musica.edel.univ-poitiers.fr/443/textus-et-musica/index.php?id=2064>.
- CHRÉTIEN DE TROYES (1975). *Le Conte du Graal (Perceval)*. Sous la dir. de Félix LECOY. T. 2. 6 t. Paris : Honoré Champion.
- (1994). *Œuvres complètes*. Sous la dir. de Daniel POIRION, Peter F. DEMBOWSKI, Philippe WALTER, Karl D. UTTI et Anne BERTHELOT. Trad. par Peter F. DEMBOWSKI, Philippe WALTER, Daniel POIRION et Anne BERTHELOT. Paris : Gallimard.
- (2009a). *Conte du Graal (Perceval)*. Sous la dir. de Pierre KUNSTMANN. Ottawa/Nancy : Université d'Ottawa, Laboratoire de Français Ancien/ATILF. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 06-05-2009. URL : <http://catalog.bfm-corpus.org/PercevalKu>.
- (2009b). *Erec et Enide*. Sous la dir. de Pierre KUNSTMANN. Ottawa/Nancy : Université d'Ottawa, Laboratoire de Français Ancien/ATILF. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 06-05-2009. URL : <http://catalog.bfm-corpus.org/ErecKu>.
- CLÉDAT, Léon (1917). *Manuel de phonétique et de morphologie historique du français*. Paris : Hachette.
- CLEMENTS, George et Samuel KEYSER (1983). *CV Phonology. A Generative Theory of the Syllable*. Cambridge, Mass. : MIT Press.
- COMBETTES, Bernard (2020). « Conjonctions de subordination et locutions conjonctives ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 897-911.
- COMBETTES, Bernard, Christiane MARCHELLO-NIZIA, Sophie PRÉVOST et Magali ROUQUIER (2020). « Syntaxe de la phrase simple ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 1220-1337.
- CORNULIER, Benoît de (1981). « H-aspirée et la syllabation des expressions disjonctives ». In : *Phonology in the 80's*. Sous la dir. de Didier L. GOYVAERTS. Gand : Story-Scientia, p. 183-230.

- CORNULIER, Benoît de (2012). « Si le mètre m'était compté... Sur la notion fallacieuse de mesure du vers ». In : *Grammaire, lexique, référence, regards sur le sens, Mélanges offerts à Georges Kleiber pour ses quarante ans de carrière*. Sous la dir. de Louis de SAUSSURE, Andrée BORILLO et Marcel VUILLAUME. Bern/Berlin/Bruxelles : Peter Lang, p. 355-377.
- CURTUS, Ernst Robert, éd. (1911). *Li quatre livre des Reis*. Dresden : Droz.
- DEES, Anthonij (1971). *Étude sur l'évolution des démonstratifs en ancien et en moyen français*. Groningen : Wolters-Noordhoff Publishing.
- (1980). *Atlas des formes et constructions des chartes françaises du 13^e siècle*. Tübingen : M. Neimeyer Verlag.
 - (1985). « Dialectes et scriptae à l'époque de l'ancien français ». In : *Revue de linguistique romane* 49, p. 87-117. DOI : [10.5169/seals-399769](https://doi.org/10.5169/seals-399769).
 - (1987). *Atlas des formes linguistiques des textes littéraires de l'ancien français*. Tübingen : M. Neimeyer Verlag.
- DEIMIER, Pierre de (1610). *L'Académie de l'art poétique*. Paris : J. de Bordeaux.
- DELL, François (1973). *Les règles et les sons : introduction à la phonologie générative*. 1^{ère} édition. Paris : Hermann.
- (1985). *Les règles et les sons : introduction à la phonologie générative*. 2^{ème} édition revue et augmentée. Paris : Hermann.
 - (2011). « Singing in Tashlhiyt Berber, a Language that Allows Vowel-Less Syllables ». In : *Handbook of the Syllable*. Leiden : Brill.
 - (2015). « Text-to-tune alignment and lineation in traditional French songs ». In : *Text and Tune. On the Association of Music and Lyrics in Sung Verse*. Sous la dir. de Teresa PROTO, Paolo CANETTIERI et Gianluca VALENTI. Frankfurt : Peter Lang, p. 183-234.
- DELL, François et Romain BENINI (2021). *La Concordance chez Racine. Rapports entre structure grammaticale et structure métrique dans le théâtre de Racine*. Paris : Classique Garnier.
- DELL, François et John HALLE (2009). « Comparing musical textsetting in French and English songs ». In : *Towards a Typology of poetic forms, from language to metrics and beyond*. Sous la dir. de Jean-Louis AROUÏ et Andy ARLEO. Amsterdam/Philadelphie : John Benjamins.
- DEMBOWSKI, Peter F., éd. (1969). *Ami et Amile, Chanson de geste*. Paris : Honoré Champion.
- DENIS PIRAMUS (1935). *La vie seint Edmund le Rei : poème anglo-normand du xii^e siècle*. Sous la dir. d'Hilding KJELLMAN. Göteborg : Elanders Boktryckeri Aktiebolag.
- DENNERY, Annie (1982). « Les notations musicales au moyen âge ». In : *Médiévales* 1, p. 89-103. DOI : <https://doi.org/10.3406/medi.1982.886>.
- DENSUSIANU, Ovid, éd. (1965 [1896]). *La prise de Cordres et de Seville : chanson de geste du xii^e siècle publiée d'après le manuscrit unique de la Bibliothèque nationale*. Réimpression de l'édition de 1896 (Paris : Firmin Didot). New York : Johnson Reprint Co.
- Dictionnaire du Moyen Français (DMF)* (2020). ATILF - CNRS & Université de Lorraine. URL : <http://www.atilf.fr/dmf>.
- Dictionnaire Électronique de Chrétien de Troyes (DÉCT)* (2015). Ottawa / Nancy : LFA, Université d'Ottawa / ATILF, CNRS & Université de Lorraine. URL : <http://www.atilf.fr/dect>.
- Dictionnaire Étymologique de l'Ancien Français (DEAF)* (1974-). Max Niemeyer, Tübingen / De Gruyter, Berlin.
- Dictionnaire Étymologique de l'Ancien Français, version électronique (DEAFél)* (2010-). Académie des Sciences de Heidelberg (Heidelberger Akademie der Wissenschaften). URL : <http://www.deaf-page.de/fr/index.php>.
- Dictionnaire Étymologique Roman (DÉRom)* (2022). ATILF - CNRS & Université de Lorraine / Université de la Sarre. URL : <http://www.atilf.fr/DERom>.
- DIVJAK, Dagmar (2019). *Frequency in Language: Memory, Attention and Learning*. Cambridge/New York : Cambridge University Press.
- DRAGONETTI, Roger (1960). *La technique poétique des trouvères dans la chanson courtoise : contribution à l'étude de la rhétorique médiévale*. Bruges : De Tempel.
- DU CANGE, Charles du Tresne (Sieur de) (1954 [1883-1887]). *Glossarium mediae et infimae latinitatis*. Sous la dir. de Leopold FAVRE. 5 t. Graz : Ak. Druk. Réimpr. de l'éd. de 1883-1887 ; première éd. 1678.
- DUFOURNET, Jean (1984). *Aucassin et Nicolette*. Paris : Flammarion.
- ENCREVÉ, Pierre (1983). « La liaison sans enchaînement ». In : *Actes de la Recherche en Sciences Sociales* 46, p. 39-66.
- (1988). *La liaison avec et sans enchaînement. Phonologie tridimensionnelle et usages du français*. Paris : Seuil.
- ENGBERT, Annick (2015). *Phonétique historique et histoire de la langue*. Louvain-la-Neuve : De Boeck Supérieur.

- ERNOUT, Alfred et Alfred MEILLET (2001 [1932]). *Dictionnaire étymologique de la langue latine*. 4^e éd. Paris : Klincksieck.
- ESTIENNE, Robert (1549). *Dictionnaire Francois latin, autrement dict Les mots Francois, avec les manieres duser diceulx, tournez en latin, corrigé et augmenté*. 2^e éd. Paris : Robert Estienne.
- (1569 [1557]). *Traicté de la grāmaire Francoise*. 2^e éd. Paris : Robert Estienne.
- ESTIENNE DE FOGIERES (1979). *Le Livre des manières*. Sous la dir. de R. Anthony LODGE. Genève : Droz.
- FABRI, Pierre (1534 [1521]). *Le Grant et vray art de pleïne rethorique*. 2^e éd. 2 t. Paris : Denis Janot, [première édition Rouen : Simon Druel]. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k1048517r>.
- FAGARD, Benjamin (2020). « Prépositions et locutions prépositives ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 856-885.
- FARAL, Edmond, éd. (1922). *Courtois d'Arras, jeu du XIII^e s.* 2nd édition, revue ; première édition 1911. Paris : Honoré Champion.
- FAUST, Noam (2023a). « Hiatus hinders syncope in Modern Hebrew ». In : *Radical* 3, p. 203-237.
- (2023b). « Metrically-conditioned syncope in Strict CV metrics ». In : *Glossa* 8, p. 1-29.
- (2023c). « The pronominal system of Saint-Pierre-le-Bost ». In : *Isogloss* 9, p. 1-16.
- FAUST, Noam et Nicola LAMPITELLI (2017). « Virtual length and the two i's of Qaraqosh Neo-Aramic ». In : *Journal of Semitic Studies* 65, p. 35-60.
- FAUST, Noam, Nicola LAMPITELLI et Shanti ULFSBJORNINN (2018). « Articles of Italian unite! Italian definite articles without allomorphy ». In : *Canadian Journal of Linguistics* 3.63, p. 1-27.
- FAY, Percival Bradshaw et John L. GRIGSBY, éd. (1972). *Jouffroi de Poitiers, roman d'aventures du XIII^e s. siècle*. Genève/Paris : Droz/Minard.
- FÉRAUD, Olivier (2016). « La Chanson de Geste, Pistes pour une interprétation musicale ». masterclass au Conservatoire National Supérieur de Musique et de Danse de Lyon, 27/04/2016.
- FLUTRE, Louis-Fernand et Cornelis SNEIJERS DE VOGEL, éd. (1935). *Li Fet des Romains compilé ensemble de Saluste et de Suetoine et de Lucan*. 2 t. Liège/Paris/Groningue : impr. G. Thone/E. Droz/J.-B. Wolters.
- FOERSTER, Wendelin, éd. (1877). *Li chevaliers as deus espees*. Halle : Niemeyer.
- (1904). « Randglossen zur Cantefable ». In : *Zeitschrift für romanische Philologie* 28, p. 492-512.
- FOUCHÉ, Pierre (1956). *Traité de prononciation française*. Paris : Klincksieck.
- (1966). *Phonétique historique du français — Les consonnes*. 2^e éd. T. 3. 3 t. Paris : Klincksieck.
- (1969 [1958]). *Phonétique historique du français — Les voyelles*. 2^e éd. T. 2. 3 t. Paris : Klincksieck.
- FOULET, Lucien (1924). « L'accent tonique et l'ordre des mots, formes faibles du pronom personnel après le verbe ». In : *Romania* 50.197, p. 55-93. URL : <https://doi.org/10.3406/roma.1924.4573>.
- GAFFIOT, Felix (1934). *Dictionnaire Latin Français*. Paris : Hachette.
- GAUTIER, Léon (1872). *La Chanson de Roland*. 2 t. Tours : Alfred Mame et Fils. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k102947j> (t. i), <https://gallica.bnf.fr/ark:/12148/bpt6k102948x> (t. ii).
- GAUTIER DE COINCI (1966). *Les Miracles de Notre-Dame. Prologues. Chansons du premier cycle. Miracle de Theophile*. Sous la dir. de Victor Frédéric KOENIG. T. 1. Paris : Droz.
- GEOFFROY DE PARIS (1956). *La chronique métrique attribuée à Geffroy de Paris*. Sous la dir. d'Armel Hugh DIVERRES. Paris : Publications de la Faculté des Lettres de l'Université de Strasbourg/Les Belles Lettres.
- GLEÛGEN, Martin-D. et Xavier GOUVERT (2007). « La base textuelle du *Nouveau Corpus d'Amsterdam* : Ancre diasystématique et évaluation philologique ». In : *Le Nouveau Corpus d'Amsterdam : actes de l'atelier de Lauterbad*. Sous la dir. de Pierre KUNSTMANN et Achim STEIN. Stuttgart : Steiner, p. 51-84.
- GLEÛGEN, Martin-D. et Claire VACHON (2013). « L'étude philologique et scriptologique du *Nouveau Corpus d'Amsterdam* ». In : *Actas del XXVI Congreso Internacional de Lingüística y Filología Románica: Valencia 2010*. Sous la dir. d'Emili CASANOVA et Cesáreo CALVO. T. 7. Berlin/Boston : De Gruyter, 191-204 [4125-4137].
- GESS, Randall, Haïke JACOBS et Bernard LAKS (2020). « Phonétique Historique — évolution depuis l'ancien français ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 450-490.
- GILLIÉRON, Jules et Edmond EDMONT (1902-1910). *Atlas linguistique de la France*. Paris : Champion.
- GODEFROY, Frédéric (1895-1902). *Complément au Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle*. 3 t. Paris : Bouillon.

- GODEFROY, Frédéric (1891-1902). *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle*. 9 t. Paris : F. Vieweg.
- GOLDBERG, Adele E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago : University of Chicago Press.
- (2003). « Constructions: A new theoretical approach to language ». In : *Trends in Cognitive Science* 7, p. 219-224.
- GOLDSMITH, John A. (1972). *Autosegmental Phonology*. Cambridge, Mass. : Thèse de doctorat, MIT.
- GOSSEN, Charles Théodore (1951). *Petite grammaire de l'ancien picard*. Paris : Klincksiek.
- (1962). « De l'Histoire des langues écrites régionales du domaine d'oïl ». In : *Revue de linguistique romane* 26, p. 271-284. DOI : <https://doi.org/10.5169/seals-399298>.
 - (1967). *Französische Skriptastudien : Untersuchungen zu den nordfranzösischen Urkundensprachen des Mittelalters*. Vienne : Hermann Böhlau Nachf.
 - (1970). *Grammaire de l'ancien picard*. Réédition (« refonte presque totale », p. 10) de GOSSEN 1951. Paris : Klincksiek.
- GOYENS, Michèle (1994). *Émergence et évolution du syntagme nominal en français*. Berne : Peter Lang.
- GRÉCO, Gérard, André CHARBONNET, Mark DE WILDE et Bernard MARÉCHA (2020). *Bailly 2020 Hugo Chávez*. éd. électronique de BAILLY et EGGER (1935). URL : <https://bailly.app>.
- GROSSEL, Marie-Geneviève (2016). « Madeleine Tyssens, *Le Chansonnier français U, publié d'après le manuscrit Paris, BNF, fr. 20050* (tome 1) [compte-rendu] ». In : *Perspectives médiévales* [En ligne] 37. URL : <http://peme.revues.org/11396>.
- GUERNES DE PONT-SAINTE-MAXENCE (1936). *Vie de saint Thomas Becket*. Sous la dir. d'Emmanuel WALBERG. Paris : Champion. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 26/01/2016. URL : <http://txm.ish-lyon.cnrs.fr/bfm/catalog/becket>.
- GUILLAUME DE LORRIS (1934). *Le roman de la rose*. Sous la dir. de Félix LECOY. T. 1. 3 t. Paris : Honoré Champion.
- GUILLAUME LE CLERC DE NORMANDIE (1892). *Le bestiaire : Das Thierbuch des normannischen Dichters Guillaume le Clerc*. Sous la dir. de Robert REINSCH. Leipzig : O. R. Reisland.
- (1973). *Le Besant de Dieu de Guillaume le Clerc de Normandie*. Sous la dir. de Pierre RUELLE. Bruxelles : éditions de l'Université de Bruxelles.
- GUILLOT-BARBANCE, Céline, Serge HEIDEN et Alexei LAVRENTIEV (2017). « Base de français médiéval : une base de référence de sources médiévales ouverte et libre au service de la communauté scientifique ». In : *Diachroniques* 7, p. 168-184.
- GUISCHART DE BEAULIU (1909). *Le Sermon de Guischart de Beauliu*. Sous la dir. d'Arvid GABRIELSON. Uppsala : C. J. Lundström.
- GUSSMAN, Edmund et Jonathan KAYE (1993). « Polish notes from a Dubrovnik Café: I. The yers. » In : *SOAS Working Papers in Linguistics and Phonetics* 3, p. 427-462.
- HALLE, John et Fred LERDAHL (1993). « A Generative Textsetting Model ». In : *Current Musicology* 55, p. 3-23.
- HALLE, Morris et Samuel Jay KEYSER (1966). « Chaucer and the study of prosody ». In : *College English* 28.3, p. 187-219.
- (1971). « Illustration and defense of a theory of the iambic pentameter ». In : *College English* 33.2, p. 154-176.
- HANSEN, Maj-Britt Mosegaard (2020). « La négation de proposition ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 1679-1694.
- HASENORH, Geneviève (1993). *Introduction à l'ancien français*. 2^e éd. Paris : SEDES.
- HAUST, Jean, éd. (1941). *Médecinaire Liègeois du XIII^e siècle et médecin naurois du XV^e, manuscrits 815 et 2769 de Darmstadt*. Bruxelles/Liège : Palais des académies/H. Vaillant-Carmanne.
- HAYES, Bruce et Abigail KAUN (1996). « The role of phonological phrasing in sung and chanted verse ». In : *The Linguistic Review* 13, p. 243-303.
- HEIDEN, Serge, Jean-Philippe MAGUÉ et Bénédicte PINCEMIN (juin 2010). « TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement ». In : *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*. Sous la dir. de Sergio BOLASCO, Isabella CHIARI et Luca GIULIANO. T. 2. 3. Rome, Italy : Edizioni Universitarie di Lettere Economia Diritto, p. 1021-1032. URL : <https://halshs.archives-ouvertes.fr/halshs-00549779>.
- HILKA, Alfons, éd. (1912-1916). *Li Romanz d'Athis et Prophelias (l'estoire d'Athenes)*. 2 t. Dresden : Niemeyer.

- HUON DE MERY (1976). *Le torneiment Anticrist by Huon de Meri. A critical edition*. Sous la dir. de Margaret O. BENDER. 2nd édition, revue et augmentée; première édition 1971. [ville non indiquée] : University Mississippi.
- JACOBS, Haike (1993). « The phonology of enclisis and proclisis in Gallo-Romance and Old French. Selected Papers from LSRL XXI ». In : *Linguistic Perspectives on Romance Languages*. T. 1. 2 t. Amsterdam : John Benjamins Publishing Company, p. 149-163.
- JACQUES BRETTEL (1932). *Le tournoi de Chauvency*. Sous la dir. de Maurice DELBOUILLE. Liège/Paris : Imp. H. Vaillant-Carmanne/Droz.
- JACQUES D'AMIENS (1868). *L'art d'amors und Li remedies d'amors. Zwei altfranzösische Lehrgedichte von Jacques d'Amiens*. Sous la dir. de Gustav KÖRTING. Leipzig : Verlag von F. C. W. Vogel.
- JACQUES DE BAISIEUX (1973). *L'œuvre de Jacques de Baisieux*. Sous la dir. de Patrick A. THOMAS. La Hague / Paris : Mouton.
- JAKOBSON, Roman (1960). « Linguistics and poetics ». In : *Style in Language*. Sous la dir. de Thomas A. SEBEOK. Cambridge (Mass.) : The MIT Press, p. 350-377.
- (1969 [1922]). *O česhskom stikhe, preimuščestvenno v sopostaolenii s russkim*. Providence : Brown University Press.
- JEAN DE MEUN (1897). *L'art de chevalerie, traduction du de re militari de Végèce*. Sous la dir. d'Ulysse ROBERT. Paris : Firmin Didot et C^{ie}.
- (1934). *Traduction de la première épître de P. Abélard*. Sous la dir. de Charlotte CHARRIER. Paris : Honoré Champion.
- JEAN LE DIACRE (1883). « La Vie de saint Grégoire le Grand traduite du latin par frère Angier, religieux de Sainte-Frideswide ». In : *Romania* 12.46-47, p. 145-208. URL : https://www.persee.fr/doc/roma_0035-8029_1883_num_12_46_6260.
- JEAN LE MARCHANZ (1973). *Miracles de Notre Dame de Chartres*. Sous la dir. de Pierre KUNSTMANN. Ottawa/-Chartres : Éditions de l'Université d'Ottawa/Société archéologique d'Eure-et-Loir.
- JEAN PRIORAT DE BESANÇON (1897). *Li abrejance de l'ordre de chevalerie / mise en vers de la traduction de Végèce de Jean de Meun par Jean Priorat*. Sous la dir. d'Ulysse ROBERT. Paris : Firmin Didot.
- JEAN RENART (1913). *Le Lai de l'Ombre*. Sous la dir. de Joseph BÉDIER. Paris : Firmin Didot. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k5180f>.
- JEHAN BODEL (1965). *Le Romand de Cassidorus*. Sous la dir. d'Albert HENRY. 2nd édition revue; première édition 1962, Paris : PUF. Bruxelles : Presses universitaires de Bruxelles.
- JEHAN MALKARAUME (1987). *La bible de Jehan Malkaraume*. Sous la dir. de Jean Robert SMITH. 2 t. Assen : van Gorcum.
- JOHNSTON, Ronald Carlyle et Douglas David Roy OWEN, éd. (1972). *Two old French Gauvain romances*. Edinburgh/London : Scottish academic press.
- JOHAM DE LA CHAPELE DE BLOIS (1932). *Le conte dou barril*. Sous la dir. de R. Chapman BATES. London/Oxford : Yale University Press.
- KAYE, Jonathan (1990). « 'Coda' licensing ». In : *Phonology* 2.7, p. 301-330.
- KENNEDY, Elspeth, éd. (1980). *Lancelot do lac: the non-cyclic old french prose romance*. 2 t. Oxford : Clarendon Press.
- KENSTOWICZ, Michael et Jerzy RUBACH (1987). « The Phonology of Syllabic Nuclei in Slovak ». In : *Language* 63, p. 463-497.
- KIPARSKY, Paul (2006). « A modular metrics for folk verse ». In : *Formal Approaches to Poetry: Recent Developments in Metrics*. Sous la dir. de B. Elan DRESHER et Nila FRIEDBERG. Berlin / New York : Mouton de Gruyter, p. 7-49.
- KIPARSKY, Paul (1977). « The Rhythmic Structure of English Verse ». In : *Linguistic Inquiry* 2.8, p. 189-247.
- KUKENHEIM, Louis (1971). « Rôle de la prosodie dans l'histoire de la langue française ». In : *Mélanges de philologie romane dédiés à la mémoire de Jean Boutière*. T. 1. 2 t. Liège : Soledi, p. 317-331.
- KUNSTMANN, Pierre et Achim STEIN (2007). « Le Nouveau Corpus d'Amsterdam ». In : *Le Nouveau Corpus d'Amsterdam : actes de l'atelier de Lauterbad*. Sous la dir. de Pierre KUNSTMANN et Achim STEIN. Stuttgart : Steiner, p. 9-28.
- LA CHAUSSÉE, François de (1989 [1974]). *Initiation à la phonétique historique de l'ancien français*. 3^e édition revue et augmentée. Lille, Paris : Klincksieck.

- LABOV, William (1982). « Building on empirical foundations ». In : *Perspectives on Historical Linguistics*. Sous la dir. de Winfred P. LEHMANN et Yakov MALKIEL. Amsterdam/Philadelphia : John Benjamins, p. 17-92.
- LÅNGFORS, Arthur (1921). « Le miroir de vie et de mort par Robert de l'Omme (1266), modèle d'une moralité wallonne du xv^e s. (premier article) ». In : *Romania* 47.188, p. 511-531. doi : <https://doi.org/10.3406/roma.1921.4453>.
- (1924). « Le miroir de vie et de mort par Robert de l'Omme (1266), modèle d'une moralité wallonne du xv^e s. (deuxième et dernier article) ». In : *Romania* 50.197, p. 14-53. doi : <https://doi.org/10.3406/roma.1924.4572>.
- LANGLOIS, Ernest (1925). *Le couronnement de Louis, chanson de geste du xii^e s.* 2nd édition, revue. Première édition : 1920. Paris : Honoré Champion.
- LAVRENTIEV, Alexei, Céline GUILLOT-BARBANCE et Thomas RAINSFORD, éd. (2018). *Séquence de sainte Eulalie*. Lyon : ENS de Lyon. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 31-7-2014. URL : <http://catalog.bfm-corpus.org/eulaliBfm>.
- LAVRENTIEV, Alexei, Thomas RAINSFORD et Céline GUILLOT-BARBANCE, éd. (2014). *Serments de Strasbourg*. Lyon : ENS de Lyon. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 31-7-2014. URL : <http://catalog.bfm-corpus.org/strasbBfm>.
- LEMÂITRE, Henry et Gaston RAYNAUD, éd. (1914). *Le Roman de Renart le contrefait*. T. 2. Paris : Champion.
- LERDAHL, Fred et Ray JACKENDOFF (1987). *A Generative Theory of Tonal Music*. Cambridge (Mass.) : MIT Press.
- LLAMAS-POMBO, Elena (2020). « Ponctuation ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 592-614.
- LODGE, Anthony (1997). *Le français, histoire d'un dialecte devenu langue*. Trad. par Cyril VEKEN. Paris : Fayard.
- (2008). « Les débuts de la standardisation du français ». In : *SHS Web Conf. : Congrès Mondial de Linguistique Française - CMLF 2008*, p. 367-376. doi : [10.1051/cmlf08320](https://doi.org/10.1051/cmlf08320).
- LOTE, Georges (1949-1996 [1991]). *Histoire du vers français*. Paris / Aix-en-Provence : Bovin et Hatier / Presses de l'Université de Provence, [éd. num. Aix-en-Provence : Presses Universitaires de Provence]. URL : <https://books.openedition.org/author?name=lote+georges>.
- LOWENSTAMM, Jean (1996). « CV as the only syllable type ». In : *Current trends in Phonology. Models and Methods*. Sous la dir. de Jacques DURAND et Bernard LAKS. Salford, Manchester : ESRI, p. 419-441.
- LUQUIENS, Frederick Bliss (1919 [1909]). *An introduction to old french phonology and morphology*. 2^e éd. New Haven : Yale University Press.
- MACÉ DE LA CHARITÉ (1977). *La Bible de Macé de la Charité*. Sous la dir. de Paul E. R. VERHUYCK. T. 2. Leiden : Presse Universitaire.
- MALHERBE, François de ([1862]). *Œuvres*. Sous la dir. de M. L. LALANNE. Paris : Hachette, [url t. iv]. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k5210r>.
- MALMBERG, B., éd. (1940). *Roman du Comte de Poitiers*. Lund / Copenhague : Gleerup / Munksgaard. DEAF : [ComtePoitM](https://doi.org/10.1017/9780804720000).
- MANDACH, André de, éd. (1970). *Chronique dite Saintongeaise, texte franco-occitan inédit*. Tübingen : M. Niemeyer.
- MARCHELLO-NIZIA, Christiane (1999a). *L'Évolution du français, ordre des mots, démonstratifs, accent tonique*. Paris : Armand Colin.
- (1999b). *Le français en diachronie : douze siècles d'évolution*. Paris : Orphys.
- (2015). « De JE à J' en français : étapes vers l'élision, interactions entre phonétique et syntaxe ». In : *Diachroniques* 5, p. 17-43.
- (2020a). « Les adverbes ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 912-941.
- (2020b). « Les possessifs — déterminants, adjectifs et pronoms ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 702-715.
- (2020c). « Les pronoms personnels, et EN, Y et ON ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 674-692.

- MARCHELLO-NIZIA, Christiane (2020d). « Les relatifs, interrogatifs, exclamatifs : déterminants, pronoms et adverbes ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 732-744.
- MARCHELLO-NIZIA, Christiane, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER (2020). *Grande Grammaire Historique du Français*. Berlin/Boston : Mouton de Gruyter.
- MARCHELLO-NIZIA, Christiane et Sophie PRÉVOST (2020). « Expression et position des constituants majeurs dans les divers types de propositions ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHEER. Berlin/Boston : Mouton de Gruyter, p. 1055-1219.
- MARIE DE FRANCE (1947). *Lais*. Sous la dir. d'A. EWERT. Lyon : Blackwell. DEAF : [MarieLaisE](#).
- (1958). *Le lai de Lanval*. Sous la dir. de Jean RYCHNER. Genève : Droz.
- MATTE, Édouard-Joseph (1982). *Histoire des modes phonétiques du français*. Genève : Droz.
- MAZALEYRAT, Jean (1990 [1965]). *Éléments de métrique française*. 7^e éd. Paris : Armand Colin.
- MEIGRET, LOUIS (1548). *Le menteur, ou l'incrédule de Lucian traduit de Grec en François par Louis Meigrêt Lionoës, avec une écriture q'adrant a la prolaçon Françoisze : & les rezons*. Paris : Chrestien Wechel, [éd. num. Olivier Bettens 2017]. URL : <https://virga.org/phon16/index.php?item=3>.
- (1550a). *La Réponse de Louís Meigrêt a l'Apologie de Iâques Pelletier*. Paris : Chrestien Wechel. URL : <https://gallica.bnf.fr/ark:/12148/btv1b8624658m>.
- (1550b). *Le trehtë de la gramme française*. Paris : Chrestien Wechel. URL : <https://gallica.bnf.fr/ark:/12148/btv1b8624665r>.
- (1545 [1542]). *Traité touchant le commun usage de l'écriture française*. 2^e éd. Paris : Jeanne de Marnef, veuve de Denis Janot [première édition Paris : Denis Janot. URL : <https://gallica.bnf.fr/ark:/12148/btv1b86261610>].
- MÉNARD, Philippe (1997). « Remarques sur certains emplois de com(me) en ancien français ». In : *Les formes du sens. Études de linguistique française, médiévale et générale offertes à Robert Martin à l'occasion de ses 60 ans*. Sous la dir. de Georges KLEIBER. Louvain-la-Neuve : De Boeck Supérieur, p. 257-267. DOI : [10.3917/dbu.kleib.1996.01.0257](https://doi.org/10.3917/dbu.kleib.1996.01.0257).
- (1988 [1973]). *Syntaxe de l'ancien français*. 3^e édition, revue et augmentée. Bordeaux : Bière.
- MEYER, Paul et Auguste LONGNON, éd. (1882). *Raoul de Cambrai : chanson de geste*. Paris : Firmin Didot.
- MEYER-LÜBKE, Wilhelm (1890). *Grammaire des langues romanes*. Trad. par Eugène RABIER. T. 1. Paris : E. Welter.
- (1910). « Aucassin und Nicolette ». In : *Zeitschrift für romanische Philologie* 34, p. 513-522.
- (1911-1920). *Romanisches Etymologisches Wörterbuch*. Heidelberg : Winter.
- MICHA, Alexandre (1980). *Étude sur le « Merlin » de Robert de Boron, Roman du XIII^e siècle*. Genève : Droz.
- éd. (1978-1983). *Lancelot : roman en prose du XIII^e s.* T. 7. 9 t. Genève : Droz.
- MICHEL, Francisque, éd. (1860). *Psautier d'Oxford*. Oxford : Oxford University Press. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 01/11/2005. URL : <http://txm.ish-lyon.cnrs.fr/bfm/catalog/oxfps>.
- éd. (1876). *Le Livre des psaumes*. Paris : Imprimerie Nationale.
- MILLS, Leonard R., éd. (1973). *L'Histoire de Barlaam et Josaphat, version champenoise d'après le ms. Reg. lat. 660 de la Bibliothèque Apostolique Vaticane*. Genève : Droz.
- MOIGNET, Gérard (1965). *Le pronom personnel français. Essai de psycho-systématique historique*. Paris : Klincksieck.
- éd. (1972a). *Chanson de Roland*. Paris : Bordas.
- éd. (1972b). *Chanson de Roland*. Paris : Bordas. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 24-2-2014. URL : <http://catalog.bfm-corpus.org/roland>.
- (2002 [1976]). *Grammaire de l'ancien français*. 2^e éd. Paris : Klincksieck.
- MORÉE, Cécile de et Timothée PREMAT (2021). « Multilingualism in Music: A Multidisciplinary Analysis of the Codex Reina and its Franco-Dutch repertoire (c. 1400) ». Présentation au *Utrecht seminar in Multilingual Literary Cultures in the Middle Ages*, Univ. Utrecht, 01/07/2021.
- MORIN, Yves Charles (1974). « Règles phonologiques à domaine indéterminé : chute du cheva en français ». In : *Cahier de linguistique* 4, p. 69-88.
- (2000). « La variation dialectale et l'interdiction des suites *Voyelle + e muet* dans la poésie classique ». In : *Le vers français — histoire, théorie, esthétique*. Sous la dir. de Michel MURAT. Paris : Honoré Champion, p. 193-227.

- MORIN, Yves Charles (2005). « La liaison relève-t-elle d'une tendance à éviter les hiatus ? Réflexions sur son évolution historique ». In : *Langages* 158.2, p. 8-23.
- NESPOR, Marina et Irene VOGEL (2007 [1986]). *Prosodic Phonology*. 2^e éd. Berlin : Mouton de Gruyter.
- NGUYEN, Noël, Sophie WAUQUIER-GRAVELINE, Leonardo LANCIA et Betty TULLER (2007). « Detection of liaison consonants in speech processing in French: Experimental data and theoretical implications ». In : *Segmental and Prosodic Issues in Romance Phonology*. Sous la dir. de Pilar PRIETO, Joan MASCARÓ et Maria-Josep SOLÉ. Amsterdam/Philadelphia : John Benjamins, p. 3-23.
- NIEDZIELSKI, Henri, éd. (1966). *Le roman de Helcanus*. Genève : Droz.
- NOOMEN, Willem et Nico VAN DEN BOOGAARD (1983-1998). *Nouveau recueil complet des fabliaux*. 10 t. Assen : Van Gorcum.
- NORMAND, Jacques et Gaston RAYNAUD, éd. (1877). *Aiol, chanson de geste*. Paris : Didot. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k50136>.
- NOYER, Rolf (2002). « Generative metrics and Old French octosyllabic verse ». In : *Language Variation and Change* 14, p. 119-171.
- NYROP, Kristoffer (1899). *Grammaire historique de la langue française*. T. 1. 5 t. Paris : Picard.
- PAGLIANO, Claudine (2003). *L'épenthèse consonantique en français. Ce que la syntaxe, la sémantique et la morphologie peuvent faire à la phonologie*. Thèse de doctorat, Université de Nice Sophia Antipolis.
- PALERMO, Joseph, éd. (1963). *Le Romand de Cassidorus*. T. 1. 2 t. Paris : A. et J. Picard et C^{ie}.
- PALSGRAVE, Jehan (1530). *L'Esclaircissement de la langue françoise*. London : John Haukyns.
- PARIS, Gaston (1872). « La Vie de saint Léger, texte revu sur le ms. de Clermont-Ferrand ». In : *Romania* 1.3, p. 273-302.
- (1879). « Aucassin und Nicolete (Suchier) / Aucassin et Nicolette (Bida et Paris) [compte rendu] ». In : *Romania* 30, p. 284-293. URL : https://www.persee.fr/doc/roma_0035-8029_1879_num_8_30_6472_t1_0284_0000_2.
- PARKER, Ian (1978). « À propos de la tradition manuscrite des chansons de trouvères ». In : *Revue de Musicologie* 2, p. 181-202.
- PASQUET, Emmanuel (1888). « Sermons de carême en dialecte wallon ». In : *Mémoires couronnés et autres mémoires publiés par l'Académie royale des sciences, des lettres et des beaux-arts de Belgique* 41. URL : <https://archive.org/details/mmoirescouronnse4118unse>.
- PAUPHILET, Albert, éd. (1923). *La queste del Saint Graal : roman du XIII^e siècle*. Paris : Honoré Champion.
- PÉAN GÂTINEAU (1896). *Leben und Wunderthaten des heiligen Martin: Altfranzösisches Gedicht aus dem Anfang des XIII Jahrhunderts*. Sous la dir. de Werner SÖDERHJELM. Tübingen : Litterarischer Verein.
- PELAN, Margaret MacLean et Nicol Christopher William SPENCE, éd. (1964). *Narcissus, poème du XII^e siècle*. Paris : Les Belles Lettres.
- PÉLETIER, Jacques (1555 [1550]). *Dialogue de l'ortographe é prononciacion françoëse*. 2^e éd. 2 t. Lyon : Jean de Tournes [réédition contenant l'Apologie à Louis Meigret (pp. 5-27) ; première édition Poitiers : Jan e Enguilbert de Marnef, 1550]. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k79267d>.
- PENSOM, Roger (1989). « Form and Meaning in Aucassin et Nicolette ». In : *Littera et sensus. Essays on forma nd meaning in medieval French literature preseted to John Fox*. Sous la dir. de David TROTTER. Exeter : University of Exeter, p. 63-72.
- PICOCHÉ, Jacqueline et Christiane MARCHELLO-NIZIA (1994 [1989]). *Histoire de la langue française*. 3^e édition, revue et corrigée. Paris : Nathan.
- PIERRET, Jean-Marie (1994). *Phonétique historique du français, et Notions de phonétique générale*. Louvain-la-Neuve : Peeters.
- POERCK, Guy de, éd. (1956). *Sermon sur Jonas*. Gent : Romanica Gandensia. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 2009-06-19. URL : <http://catalog.bfm-corpus.org/jonas>.
- POGGIO, Enzo et Timothée PREMAT (2019). « Le PAM, un Programme d'Analyse Métrique pour le français médiéval ». In : *Actes des Rencontres lyonnaises des jeunes chercheurs en linguistique historique*. Sous la dir. de Timothée PREMAT et Ariane PINCHE. Diachronies contemporaines. DOI : 10.5281/zenodo.3464477. URL : <https://hal.archives-ouvertes.fr/hal-02320550>.
- POPE, Mildred Katherine (1966 [1934]). *From Latin to Modern French with Especial Consideration of Anglo-Norman*. 2^e éd. Manchester : Manchester University Press.

- POST, Brechtje (2000). *Tonal and Phrasal Structures in French Intonation*. La Haye : Thesus.
- PREMAT, Timothée (2017). *Conflicts et synergies des patrons structuraux dans la musique des trouvères, Investigations sur l'interface entre musique, métrique et langue*. Mémoire de Master, Université Jean Moulin Lyon 3.
- (2022). « Le sandhi externe en très ancien français ». présentation au congrès de Réseau Français de Phonologie (RFP), Porto, 09/02/2022.
 - (2023). « Comparalem : étudier la variation graphique dans des corpus annotés de français médiéval ». présentation à la journée d'étude *Le traitement numérique des matériaux galloromans : sources anciennes, nouveaux problèmes*, Université de Neuchâtel, 16/02/2023.
- PREMAT, Timothée et Sophie CHOUVION (2021). « Adaptations de la mélodie aux variations de longueur du vers aux XII^e et XIII^e s. » In : *Rhythmica*. Sous la dir. d'Alexandra ODDO, Jean-Claude ANSCOMBRE et Bernard DARBORD, p. 243-270. DOI : [10.5944/rhythmica.32387](https://doi.org/10.5944/rhythmica.32387).
- PREMAT, Timothée, Sophie CHOUVION et Axelle VERNER (2021). « Repetitions, rests insertion and Schwa in 16th century French Polyphony : An Emergence Sub-Grammar in Fresneau's Songs? » In : *The Emergence of Grammar : A Closer Look at Dialects between Morphology and Phonology*. Sous la dir. de Michela Russo. New York : Academic Press, p. 247-301.
- PREMAT, Timothée et ENZO POGGIO (à paraître). « Restituer des fragments de phonologie avec le PAM (*Programme d'Analyse Métrique*) : l'élision de schwa et des voyelles pleines dans les monosyllabes fonctionnels en très ancien français ». In : *Promptus*.
- PRÉVOST, Sophie, Céline GUILLOT, Alexei LAVRENTIEV et Serge HEIDEN (2013). « Jeu d'étiquettes morphosyntaxiques CATTEX2009 ». version 2.0. URL : <http://bfm.ens-lyon.fr/IMG/xml/cattex2009.xml>.
- PRÉVOST, Sophie et Achim STEIN, éd. (2013). *Syntactic Reference Corpus of Medieval French (SRCMF)*. Lyon/Stuttgart : ENS de Lyon ; Lattice, Paris ; ILR University of Stuttgart. URL : <http://srcmf.org>.
- PRICE, Glanville (1985 [1971]). *The French language: past and present*. 4^e édition revue et corrigée [1^{ère} éd. à Londres chez E. Arnold]. Londres : Grant & Cutler.
- PRINCE, Alan et Paul SMOLENSKY (2004). *Optimality Theory : Constraint interaction in generative grammar*. Malden (Mass.)/Oxford/Victoria : Blackwell publishing.
- PSEUDO-NICODÈME (1973). *L'Évangile de Nicodème : les versions courtes en ancien français et en prose*. Sous la dir. d'Alvin Earle FORD. Genève : Droz.
- RAINSFORD, Thomas (2010). « Rhythmic change in the medieval octosyllable and the development of group stress ». In : *Congrès Mondial de Linguistique Française - CMLF 2010*. DOI : [10.1051/cmlf/2010174](https://doi.org/10.1051/cmlf/2010174).
- (2011a). « Dividing lines: the changing syntax and prosody of the mid-line break in medieval French octosyllabic verse ». In : *Transactions of the Philological Society* 109.3, p. 265-283.
 - (2011b). « The Emergence of Group Stress in Medieval French ». Thèse de doct. Cambridge : Cambridge University, St. John's College.
 - (2020). « Syllable structure and prosodic words in Early Old French ». In : *Papers in Historical Phonology* 5, p. 63-89. DOI : [10.2218/pihph.5.2020.4433](https://doi.org/10.2218/pihph.5.2020.4433).
- RAINSFORD, Thomas et Christiane MARCHELLO-NIZIA, éd. (2018). *Vie de saint Alexis*. Lyon : ENS de Lyon. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 24-10-2018. URL : <http://catalog.bfm-corpus.org/AlexisRaM>.
- RAINSFORD, Thomas et Olga SCRIVNER (2014). « Metrical annotation for a verse treebank ». In : *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT 13)*. Sous la dir. de Verena HENRICH, Erhard HINRICHS, Daniël KOK, Petya OSENOVA et Adam PRZEPIÓRKOWSKI. Tübingen : Département de Linguistique, Université de Tübingen.
- RAMÉE, Pierre de la (1572). *Grammaire*. Paris : André Wechel.
- RANKKA, Erik, éd. (1982). *Li ver del juïse : sermon en vers du 12^e siècle*. 2nd édition, revue et augmentée ; première édition 1971. Uppsala : Almqvist & Wiksell.
- RAOUL DE HOUDENC (1983). *Le roman des Eles*. Sous la dir. de Keith BUSBY. Amsterdam/Philadelphia : J. Benjamins.
- REGULA, Moritz (1955). *Historische Grammatik des Französischen*. T. 1. 3 t. Heidelberg : C. Winter.
- RENAULT (1931). *La Vie de saint Jehan Bouche d'Or et la Vie de sainte Dieudonnee, sa mere*. Sous la dir. d'Hermine DIRICKX-VAN DER STRAETEN. Liège : Vaillant-Carmanne.
- RENAUT DE BEAUJEU (1929). *Le bel inconnu : roman d'aventures*. Sous la dir. de Gwladys Perrie WILLIAMS. Paris : Honoré Champion.

- REY, Alain, éd. (2012). *Dictionnaire historique de la langue française*. 3 t. Paris : Dictionnaires Le Robert.
- RIEGEL, Martin, Jean-Christophe PELLAT et Rioul RENÉ (2009 [1994]). *Grammaire méthodique du français*. 4^e édition, entièrement revue. Paris : Presses Universitaires de France.
- ROACH, William, éd. (1971). *The Continuations of the old French "Perceval" of Chrétien de Troyes: The second continuation*. T. iv. Philadelphia : The American Philosophical Society.
- ROBERT DE CLARI (1956). *Robert de Clari. La conquête de Constantinople*. Sous la dir. de Philippe LAUER. Paris : Champion. URL : <https://archive.org/details/laconquetedeconstantinoplelauer>.
- RONCARD, Pierre de (1585). *Art poetique francoys*. Paris : Guillaume Linocier. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k70021c>.
- (1866 [1565]). « Abbrégé de l'art poetique françois ». In : *Œuvres complètes*. Sous la dir. de Prosper BLANCHEMIN. Paris : A. Franck, [première édition Paris : Gabriel Buon]. URL (1565, facs.) : <https://gallica.bnf.fr/ark:/12148/bpt6k4601j>.
- ROQUES, Mario, éd. (1929a). *Aucassin et Nicolette*. Paris : Honoré Champion. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 2018-10-19. URL : <http://catalog.bfm-corporus.org/aucassin>.
- éd. (1929b). *Aucassin et Nicolette, Chantefable du XIII^e s.* Paris : Honoré Champion.
- éd. (1955). *Aucassin et Nicolette*. Paris : Champion.
- ROSENBERG, Samuel N., Margaret L. SWITTEN et Gérard LE VOT (1998). *Songs of the troubadours and trouvères: an anthology of poems and melodies*. New York/Londres : Garland.
- ROSENBERG, Samuel N., Hans TISCHLER et Marie-Geneviève GROSSEL (1995). *Chansons des trouvères, chanter m'estuet*. Paris : Librairie Générale Française.
- RUBACH, Jerzy (1986). « Abstract vowels in three dimension phonology: the yers ». In : *The Linguistic Review* 5, p. 247-280.
- SALVERDA DE GRAVE, Jean-Jacques, éd. (1964-68). *Eneas, roman du XII^e siècle*. 2 t. Paris : Honoré Champion.
- SAMPSON, Rodney (1980). « On the history of final vowels from Latin to Old French ». In : *Zeitschrift für Romanische Philologie* 96, p. 23-48.
- SANGIOVANNI, Fabio (2020). « La cesura epica come problema lirico ». In : *Textus & Musica [en ligne]* 2. URL : <https://textus-et-musica.edel.univ-poitiers.fr:443/textus-et-musica/index.php?id=783>.
- SARGANT-BAUR, Barbara Nelson et Robert Francis COOK (1981). *Aucassin et Nicolette: a critical bibliography*. Londres : Grant & Cutler.
- SCHANE, Sanford A. (1967). « L'élision et la liaison en français ». In : *Langages* 8, p. 37-59.
- (1968). « On the abstract character of French « e muet ». In : *Glossa* 2.2, p. 150-163.
- SCHAEER, Tobias (1997). « Vowel-zero alternations and their support for a theory of consonantal interaction ». In : *Certamen Phonologicum III*. Sous la dir. de Pier Marco BERTINETTO, Livio GAETA, Georgi JETCHEV et David MICHAELS. Turin : Rosenberg & Sellier, p. 76-88.
- (1998). « Governing domains are head-final ». In : *Structure and Interpretation, Studies in Phonology*. Sous la dir. d'Eugeniusz CYRAN. Lublin : Folium, p. 261-285.
- (2015). *Précis de structure syllabique*. Lyon : ENS Éditions.
- SCHAEER, Tobias, Guylaine BRUN-TRIGAUD, Laurent VANNI, Pierre-Aurélien GEORGES, Alex STEIN, Abdurrahmaan IQBAL et Timothée PREMAT (2022). *Atlas Dees Électronique* 22. version beta : Université Côte d'Azur & CNRS. URL : <http://atlasdees.unice.fr/wordpress/>.
- SCHERER, A. (1877). « Li prière Theophilus ». In : *Zeitschrift für romanische Philologie* 1, p. 247-258. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k15851g>.
- SCHWAN, Eduard (1886). *Die altfranzösischen Liederhandschriften : ihr Verhältnis, ihre Entstehung und ihre Bestimmung*. Berlin : Weidmannsche Buchhandlung.
- SÉGÉRAL, Philippe et Tobias SCHAEER (2015). « Présentation ». In : *Diachroniques* 5, p. 7-15.
- (2020). « Phonétique Historique ». In : *Grande Grammaire Historique du Français*. Sous la dir. de Christiane MARCHELLO-NIZIA, Bernard COMBETTES, Sophie PRÉVOST et Tobias SCHAEER. Berlin/Boston : Mouton de Gruyter, p. 203-449.
- SEGRE, Cesare, éd. (1971). *La Chanson de Roland*. Milan/Naples : Ricciardi.
- SELKIRK, Elisabeth (1981). « On prosodic structure and its relation to syntactic structure ». In : *Nordic Prosody II: Paper from a Symposium*. Sous la dir. de Thorstein FRETHEIM. Trondheim : TAPIR, p. 111-140.
- (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, Mass. : MIT Press.

- SELKIRK, Elisabeth (1986). « On derived domains in sentence phonology ». In : *Phonology* 3, p. 371-405.
- (1996). « The prosodic structure of function words. » In : *Signal to syntax : bootstrapping from syntax to grammar in early acquisition*. Sous la dir. de James MORGAN et Katherine DEMUTH. Mahwah, N.J. : Erlbaum, p. 187-213.
 - (2011). « The Syntax-Phonology Interface ». In : *The Handbook of Phonological Theory*. Sous la dir. de John GOLDSMITH, Jason RIGGLE et Alan YU. 2^e éd. Oxford : Blackwell Publishing, p. 435-484.
 - (1980 [1972]). *The Phrase Phonology of English and French*. New York : Garland Publishing. Thèse de doctorat, MIT, soutenue en 1972.
- SETTEGAST, FRANZ (1879). « Calendre und seine Kaiserchronik ». In : *Romanische Studien* 3, p. 65-75.
- SHORT, Ian (1973). « The Oxford Manuscript of the Chanson de Roland: A Palaeographic Note ». In : *Romania* 94.374, p. 221-231. URL : https://www.persee.fr/doc/roma_0035-8029_1973_num_94_374_2366.
- éd. (1990). *La Chanson de Roland*. Paris : Librairie Générale Française.
 - (2007). *Manual of Anglo-Norman*. Londres : Anglo-Norman Text Society.
- SMIRKE, E. (1859). « Ancient ordinances of the gild merchant of the town of Southampton ». In : *The Archeological Journal* XVI, p. 284-296.
- SOL, Hendrik Bastiaan, éd. (1977). *La Vie du pape Saint Grégoire, huit versions françaises médiévales de la légende du bon pêcheur*. Amsterdam : Rodopi.
- SPENCER, Andrew (1986). « A non-linear analysis of vowel-zero alternations in Polish ». In : *Journal of Linguistics* 22, p. 249-280.
- STEIN, Achim, Pierre KUNSTMANN et Martin-D. GLEßGEN (2006). *Nouveau Corpus d'Amsterdam. Corpus informatique de textes littéraires d'ancien français (ca 1150-1350)*. établi par Anthonij Dees (Amsterdam : 1987), remanié par Achim Stein, Pierre Kunstmann et Martin-D. Gleßgen, Stuttgart : Institut für Linguistik/Romanistik.
- STRAKA, Georges (1979). *Les sons et les mots. Choix d'études de phonétique et de linguistique*. Paris : Klincksieck.
- STUIP, René Ernst Victor, éd. (1970). *La Chastelaine de Vergi*. La Haye/Paris : Mouton.
- SUCHIER, Hermann (1879). *Reimpredigt*. Halle : M. Niemeyer.
- (1891). « Bruchstück des Romanz des Eles von Raol von Houdan. » In : *Mélanges de philologie romane dédiés à Carl WAHLUND à l'occasion du cinquantième anniversaire de sa naissance (7 janvier 1806)*. Mâcon : Protat frères, p. 29-39.
- SYLVIUS [DU BOIS], Jacques (1531). *In linguam gallicam Isagoge, una cum eiusdem grammatica latino-gallica, ex Hebræis, Græcis et Latinis authoribus*. Paris : Robert Estienne. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k50952q>.
- THAON, Philippe de (1924). *Lapidaire alphabétique*. Sous la dir. de Paul STUDER et Joan EVANS. Lyon : ENS de Lyon. Publié en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 21-10-2009. L'attribution à Philippe de Thaon est probable. URL : <http://catalog.bfm-corpus.org/Lapidal>.
- THORDSTEIN, Arvid, éd. (1941). *Le bestiaire d'amour rimé*. Lund/Copenhague : Munksgaard.
- THUROT, Charles (1881-1883). *De la prononciation française depuis le commencement du XVI^e siècle, d'après les témoignages des grammairiens*. T. 1. Paris : Imprimerie Nationale [réimprimé en 1966 par Slatkine à Genève]. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k4617p>.
- TLFi (1994). *TLFi : Trésor de la Langue Française informatisé*. ATILF — CNRS & Université de Lorraine. URL : <http://www.atilf.fr/tlfi>.
- TOBLER, Adolf et Erhard Friedrich LOMMATZSCH ([2002] 1925-1976). *Altfranzösisches Wörterbuch*. Sous la dir. d'Erhard BLUMENTHAL et Achim STEIN. Stuttgart : F. Steiner. Édition électronique, originellement distribuée sur cédérom et DVD-ROM, accessible en ligne. URL : <https://www.ling.uni-stuttgart.de/institut/ilr/toblerlommatszsch/index.htm>.
- TODD, Henry Alfred (1900). « La vie de Sainte Catherine d'Alexandrie, as Contained in the Paris Manuscript *La Clayette* ». In : *Publications of the Modern Language Association* 1.15, p. 17-73.
- TRANEL, Bernard (1987). *The Sounds of French: An introduction*. Cambridge : Cambridge University Press.
- (1995a). « Current issues in French phonology: Liaison and position theories ». In : *A Handbook of Phonological theorys*. Sous la dir. de John GOLDSMITH. Oxford : Blackwell, p. 798-816.
 - (1995b). « The representation of French final consonants and related issues ». In : *Linguistic studies in Romance languages*. Sous la dir. de Jon AMASTAE, Grant GOODALL, Marianne PHINNEY et Mario MONTALBETTI. Amsterdam/Philadelphia : John Benjamins, p. 53-78.

- TRANEL, Bernard (1996). « French liaison and elision revisited: A unified account within Optimality Theory ». In : *Romance Linguistics in Los Angeles*. Sous la dir. de Claudia PARODI, Carlos QUICOLI, Mario SALTARELLI et Maria Luisa ZUBIZARRETA. Washington : Georgetown University Press, p. 433-455.
- (2000). « Aspects de la phonologie du français et la théorie de l’optimalité ». In : *Langue française* 126, p. 39-72. DOI : [10.3406/lfr.2000.989](https://doi.org/10.3406/lfr.2000.989).
- VAUGELAS, Claude Favre de (1880 [1647]). *Remarques sur la langue française*. 2 t. Versailles/Paris : Cerf et fils/J. Baudry [réédition contenant un supplément (nouvelles remarques posthumes et observations); première édition Paris : La veuve Jean Camusat et Pierre Le Petit].
- VERLUYTEN, S. Paul (1988). « Introduction ». In : *La phonologie du schwa français*. Sous la dir. de S. Paul VERLUYTEN. Amsterdam/Philadelphia : John Benjamins Publishing Company, p. 1-15.
- VISING, Johan (1895). « Quomodo in den romanischen Sprachen ». In : *Abhandlungen herrn prof. dr. Adolf Tobler*. Halle : Max Niemeyer, p. 113-123. URL : <https://archive.org/details/abhandlungenher00unkngoog>.
- VITZ, Evelyn B. (2007). « Variegated performance of *Aucassin et Nicolette* ». In : *Cultural performances in medieval France: Essays in honor of Nancy Freeman Regalado*. Sous la dir. d’Eglal DOSS-QUINBY, Roberta L. KRUEGER et E. Jane BURNS. Woodbridge : D.S. Brewer, p. 235-245.
- WACE (1932a). *La Vie de sainte Marguerite*. Sous la dir. d’Elisabeth A. FRANCIS. Paris : Champion.
- (1932b). *La vie de Sainte Marguerite*. Sous la dir. d’Elisabeth A. FRANCIS. Paris : Champion.
- WALKER, Douglas C. (1981). *An Introduction to Old French Morphophonology*. Ottawa : Marcel Didier.
- WALLENSKÖLD, Axel, éd. (1907-1909). *Florence de Rome, chanson d’aventure du premier quart du XIII^e siècle*. 2 t. Paris : Firmin Didot.
- WALPOLE, Ronald R. (1948). « The Burgundian Translation of the ‘Pseudo-Turpin Chronicle’ in Bibliothèque Nationale (French MS 25438) ». In : *Romance Philology* 2.2, p. 177-215.
- (1949). « The Burgundian Translation of the ‘Pseudo-Turpin Chronicle’ in Bibliothèque Nationale French MS 25438 (Part II) ». In : *Romance Philology* 3.2, p. 83-116.
- WARTBURG, Walther von (1922-1967). *Französisches Etymologisches Wörterbuch*. Sous la dir. d’Hans-Erich KELLER. Bâle : R. G. Zbinden.
- (2022 [1922-1967]). *Französisches Etymologisches Wörterbuch (FEW) informatisé*. Sous la dir. d’Elisabeth BERCHTOLD, Myriam BERGERON-MAGUIRE, Jean-Paul CHAUVEAU, Yan GREUB, Laure GRÜNER, Juila ALLETSCRUBER et Dima CHIHAI. Version informatisée de WARTBURG (1922-1967). ATILF - CNRS & Université de Lorraine. URL : <http://www.atilf.fr/dmf>.
- WAUCHIER DE DENAIN (1995). *Histoire ancienne jusqu’à César, Estoires Rogier*. Sous la dir. de Marjike de VISSER-VAN TERWISG. T. 1. 2 t. Orléans : Paradigme.
- WEBER, Alfred (1877). « Zwei ungedruckte Versionen der Theophilussage ». In : *Zeitschrift für romanische Philologie* 1.3, p. 523-540.
- WICKHAM, Hadley, Romain FRANÇOIS, Lionel HENRY, Kirill MÜLLER et Davis VAUGHAN (2023). *dplyr : A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>, <https://github.com/tidyverse/dplyr>.
- WILMOTTE, Maurice (1891). « Gloses wallonnes du ms. 2640 de Darmsladt ». In : *Études romanes dédiées à Gaston Paris*. Paris : E. Bouillon, p. 239-252.
- ZINK, Gaston (1989). *Morphologie du français médiéval*. Paris : Presses Universitaires de France.
- (2013 [1986]). *Phonétique historique du français*. 2^e éd. Paris : Presses Universitaire de France.
- ZUMTHOR, Paul (1972). *Essai de poétique médiévale*. Paris : Éditions du Seuil.

Table des figures

2.1 Répartition des textes du NCA selon la datation de Dees ou, à défaut, la datation de composition	11
2.2 Répartition des textes du NCA selon la localisation de composition.	12
2.3 Carte n° 252 de l'Atlas des formes et constructions des chartes françaises du 13 ^e siècle (DEES 1980, p. 266).	15
2.4 Carte n° 373 de l'Atlas des formes linguistiques des textes littéraires de l'ancien français (DEES 1987, p. 373).	16
2.5 Distribution des textes du NCA en fonction des trois datations disponibles.	26
2.6 Répartition des textes du NCA selon les localisations regionDees	34
2.7 Répartition des textes du NCA selon les localisations codeRegional	34
2.8 Répartition des textes du NCA selon les localisations regionDees complétées	35
2.9 Nombres de textes par région selon les différents types de localisation, périodes 1 et 2	36
2.10 Nombres de textes par région selon les différents types de localisation, périodes 3 et 4	37
2.11 Répartition des textes de la BFM selon la date de composition normalisée.	42
2.12 Répartition des textes de la BFM selon la date du manuscrit normalisée.	42
2.13 Distribution des textes de la BFM par dialectes et par siècle (date de composition normalisée).	44
2.14 <i>Chanson de Roland</i> , v. 1-2, dans le format du fichier roland.xml de la BFM.	45
2.15 <i>Chronique des Ducs de Normandie</i> , v. 1-4, dans le format du fichier BenDuc1b.xml (BFM)	45
2.16 <i>Chanson de Roland</i> , v. 2, dans le format du fichier rolandox.xml du NCA	46
2.17 Phrase musicale de la <i>Bataille d'Annezin</i>	48
2.18 <i>Aucassin et Nicolette</i> , septième section	52
2.19 Incipit du <i>Gieus de Robin et de Marion</i> , ms. BFN fr 25566	54
2.20 <i>Aucassin et Nicolette</i> , septième section	55
2.21 Exemple de notation impropre au rendu d'un mode rythmique, « <i>L'on dit qu'Amors...</i> » (ms. U)	57
2.22 Exemple de notation carrée ne rendant pas un mode rythmique, « <i>Au renouvel du tens...</i> » (ms. P)	57
2.23 Exemple de notation carrée rendant un mode rythmique, « <i>Apris ai qu'en chantant plour</i> » (ms. O)	57
2.24 Non répétition graphique de la musique « <i>Bele Doette</i> » (ms. U)	60
2.25 Répétition intra-strophique sur « <i>Bele Doette</i> » (ms. U)	61
3.1 Fonctionnement simplifié de Comparalem	67
3.2 Distribution diachronique de la var. <e>/Ø en fonction du contexte subséquent, ADJFÉMSG	76
3.3 Distribution diachronique des #C/#V en fonction de la présence d'un <e>#, ADJFÉMSG	77
3.4 Distribution par lemme de la var. <e>/Ø en fonction du contexte subséquent, ADJFÉMSG #V	78
3.5 Distribution par lemme de la var. <e>/Ø en fonction du contexte subséquent, ADJFÉMSG #C	79
3.6 Diatopie des alternances <e>/Ø, ADJFÉMSG devant #C et #V	82
3.7 Diatopie de la prévalence de l'élision, ADJFÉMSG	83
3.8 Diatopie des alternances <e>/Ø, 1100-1200, ADJFÉMSG	84
3.9 Diatopie des alternances <e>/Ø, 1200-1322, ADJFÉMSG	85
3.10 Diatopie de la prévalence de l'élision, 1100-1200, ADJFÉMSG	86
3.11 Diatopie de la prévalence de l'élision, 1200-1322, ADJFÉMSG	87
3.12 Vue synthétique du flux de travail du deuxième temps d'exécution de Comparalem	90
3.13 <i>Chanson de Roland</i> , v. 1-3, analyse du PAM	96
3.14 <i>Chanson de Roland</i> , v. 1-3, analyse du PAM	97
3.15 <i>Chanson de Roland</i> , v. 300 et 21, analyse du PAM	97
3.16 Distribution des longueurs de vers m de <i>Roland</i>	107
3.17 Distribution des longueurs de vers m de <i>Roland</i> en fonction de la position des vers dans le texte	108
3.18 Relation entre schwas, longueurs des vers et nombre des vers dans <i>Roland</i>	109
3.19 Vue synthétique du flux de travail du PAM	112
4.1 Diachronie des noms en défaut de <e>	127
4.2 Diatopie des noms en défaut de <e>	129
4.3 Distribution contextuelle des noms en défaut de <e>	130

5.1	Distribution contextuelle des ADJFÉMSG en défaut de <e>	139
5.2	Distribution par lemme de la var. <e>/Ø en fonction du contexte subséquent, ADJFÉMSG #V	140
5.3	Distribution par lemme de la var. <e>/Ø en fonction du contexte subséquent, ADJFÉMSG #C	141
5.4	Extrait de la fig. 5.2, position de psautier	143
5.5	Extrait de la fig. 5.3, position de psautier	143
5.6	Diachronie des fréquences des lemmes à défaut de <e> et de tous les lemmes, ADJFÉMSG	145
5.7	Distribution des lemmes en fonction de leur fréquence et de leurs taux de défaut de <e>, ADJFÉMSG	146
5.8	Arc d'hyperbole de fonction $y = 1/x$	146
5.9	Extrait de la fig. 5.7, tracé hyperbolique	147
5.10	Distribution contextuelle des ADJFÉMSG en défaut de <e>	154
5.11	Distribution diatopique des alternances <e>/Ø, ADJFÉMSG #C et #V	156
5.12	Distribution diatopique de la prévalence de l'élision, ADJFÉMSG	157
5.13	Diatopie des alternances <e>/Ø, 1100-1150, ADJFÉMSG	159
5.14	Diatopie de la prévalence de l'élision, 1100-1150, ADJFÉMSG	160
5.15	Diatopie des alternances <e>/Ø, 1151-1200, ADJFÉMSG	161
5.16	Diatopie de la prévalence de l'élision, 1151-1200, ADJFÉMSG	162
5.17	Diatopie des alternances <e>/Ø, 1201-1250, ADJFÉMSG	163
5.18	Diatopie de la prévalence de l'élision, 1201-1250, ADJFÉMSG	164
5.19	Diatopie des alternances <e>/Ø, 1251-1322, ADJFÉMSG	165
5.20	Diatopie de la prévalence de l'élision, 1251-1322, ADJFÉMSG	166
5.21	Distinction entre anglo-normand et dialectes continentaux, ADJFÉMSG en défaut de <e> __#V	175
5.22	Diachronie des défauts de <e>, a.n. vs a.fr.cont, ADJFÉMSG	175
6.1	<i>Roland</i> , ms. d'Oxford, fol. 4v, l. 5 (v. 203)	184
6.2	Distribution diachronique des PRÉP en défaut de <e>	187
6.3	Distribution par lemme de la var. <e>/Ø en fonction du contexte subséquent, PRÉP #V	189
6.4	Distribution par lemme de la var. <e>/Ø en fonction du contexte subséquent, PRÉP #C	190
6.5	<i>Roland</i> , ms. d'Oxford, fol. 3v, l. 8 (v. 148)	191
6.6	<i>Roland</i> , ms. d'Oxford, fol. 38r, l. 14 (v. 2089)	191
6.7	<i>Roland</i> , ms. d'Oxford, fol. 59r, l. 1 (v. 3250)	192
6.8	<i>Roland</i> , ms. d'Oxford, fol. 35v, l. 1 (v. 1941)	194
6.9	<i>Roland</i> , ms. d'Oxford, fol. 36r, l. 8 (v. 1970)	195
6.10	Diachronie des occurrences de lemme à défaut de <e> (PRÉP) et de toutes les occurrences de PRÉP	199
6.11	Distribution des lemmes en fonction de leur fréquence et de leurs taux de défaut de <e>, PRÉP	200
6.12	Distribution contextuelle des PRÉP en défaut de <e>	206
6.13	Extrait de la fig. 6.12.b, plateau de la régression locale __#V	207
6.14	Distribution contextuelle des PRÉP (sauf « (de)sor ») en défaut de <e>	209
6.15	<i>Roland</i> , ms. d'Oxford, fol. 18v, l. 24-25 (v. 1006-1007)	210
6.16	<i>Roland</i> , ms. d'Oxford, fol. 68r, l. 9 (v. 3765)	210
6.17	Distribution contextuelle des PRÉP « entre » en défaut de <e>	212
6.18	Distribution contextuelle des PRÉP « contre » et « encontre » en défaut de <e>	214
6.19	Distribution contextuelle des PRÉP « sor » et « desor » en défaut de <e>	216
6.20	Diatopie des alternances <e>/Ø, PRÉP __#C	217
6.21	Diatopie des alternances <e>/Ø, PRÉP __#V	218
6.22	Diatopie de la prévalence de l'élision, PRÉP	219
6.23	Diatopie des alternances <e>/Ø, PRÉP « de » __#V	220
6.24	Diatopie des alternances <e>/Ø, PRÉP « entre » __#V	222
6.25	Diatopie des alternances <e>/Ø, PRÉP « (de)sor » __#C	224
6.26	Diatopie des alternances <e>/Ø, PRÉP « (de)sor » __#V	225
6.27	Diatopie des alternances <e>/Ø, 1100-1200, PRÉP « de »	227
6.28	Diatopie des alternances <e>/Ø, 1201-1322, PRÉP « de »	228
6.29	Diatopie des alternances <e>/Ø, 1100-1200, PRÉP « entre »	231
6.30	Diatopie des alternances <e>/Ø, 1201-1322, PRÉP « entre »	232
6.31	Diachronie des défauts de <e>, a.n. vs a.fr.cont, PRÉP « de »	240

6.32	Diachronie des défauts de <e>, a.n. vs a.fr.cont, PRÉP « de » et « entre » __#V	241
6.33	Distinction entre anglo-normand et dialectes continentaux, PRÉP « de » en défauts de <e> __ #V	242
6.34	Diachronie des défauts de <e>, a.n. vs a.fr.cont, PRÉP « sor » et « desor »	243
6.35	Extrait (bis) de la fig. 6.12.b, plateau de la régression locale __#V	244
7.1	Diachronie des Adv en défaut de <e>	256
7.2	Diachronie des Adv (*-s) en défaut de <e>	257
7.3	Diachronie des Adv (*-s et sauf « ensemble ») en défaut de <e>	258
7.4	Distribution par lemme de la var. <e>/Ø en fonction du contexte subséquent, Adv #V	260
7.5	Distribution par lemme de la var. <e>/Ø en fonction du contexte subséquent, Adv #C	261
7.6	Diachronie des occurrences de lemme à défaut de <e> (Adv) et de toutes les occurrences de Adv	265
7.7	Distribution des lemmes en fonction de leur fréquence et de leurs taux de défaut de <e>, Adv	266
7.8	Distribution contextuelle des Adv en défaut de <e>	269
7.9	Diatopie des alternances <e>/Ø, Adv __#C	273
7.10	Diatopie des alternances <e>/Ø, Adv __#C (nb. de textes / région > 5)	274
7.11	Diatopie des alternances <e>/Ø, Adv __#V	275
7.12	Diatopie des alternances <e>/Ø, Adv __#V (nb. de textes / région > 5)	276
7.13	Diatopie de la prévalence de l'élision, Adv __#V	277
7.14	Diatopie des alternances <e>/Ø, 1100-1150, Adv __#V	279
7.15	Diatopie de la prévalence de l'élision, 1100-1150, Adv __#V	280
7.16	Diatopie des alternances <e>/Ø, 1151-1200, Adv __#V	281
7.17	Diatopie de la prévalence de l'élision, 1151-1200, Adv __#V	282
7.18	Diatopie des alternances <e>/Ø, 1201-1250, Adv __#V	284
7.19	Diatopie de la prévalence de l'élision, 1201-1250, Adv __#V	285
7.20	Diatopie des alternances <e>/Ø, 1251-1322, Adv __#V	286
7.21	Diatopie de la prévalence de l'élision, 1251-1322, Adv __#V	287
7.22	Diachronie des défauts de <e>, a.n. vs a.fr.cont, Adv	294
7.23	Diachronie des défauts de <e>, a.n. vs a.fr.cont, Adv __#V	296
8.1	Distribution contextuelle des ARTDÉFMASC CRsg en défaut de <e>	305
8.2	Distribution contextuelle des ARTDÉFFÉMCS en défaut de <V>	307
8.3	Distribution contextuelle des ARTDÉFFÉMSG (<i>la, l</i>) en défaut de <a>	308
8.4	Distribution contextuelle des ARTDÉFMASC CSsg en défaut de <i>	312
8.5	<i>Roland</i> , ms. d'Oxford, fol. 42v, pénultième l. (v. 2354)	313
8.6	Distribution contextuelle des ARTINDÉFFÉMSG en défaut de <e>	315
8.7	Diatopie des alternances <V>/Ø, ARTDÉFMASC CSsg __#V	318
8.8	Diatopie des alternances <e>/Ø, ARTINDÉFFÉMSG __#V	319
8.9	Diachronie des défauts de <e>, a.n. vs a.fr.cont, ARTINDÉFsg <i>un(e)</i> __#V	320
8.10	Distribution contextuelle des ARTDÉFMASC Spl en défaut de <i>	322
9.1	Distribution contextuelle des PROPERSMASC CR3sg en défaut de <e>	342
9.2	Diachronie des défauts de <e>, a.n. vs a.fr.cont, PROPERSMASC CR3sg	343
9.3	Diatopie des alternances <e>/Ø, PROPERSMASC CR3sg __#V	344
9.4	Diatopie des alternances <e>/Ø, PROPERSMASC CR3sg __#C	345
9.5	Distribution contextuelle des PROPERS CR2sg en défaut de <e>	347
9.6	Diachronie des défauts de <e>, a.n. vs a.fr.cont, PROPERS CR2sg	350
9.7	Diatopie des alternances <e>/Ø, PROPERS CR2sg __#C	351
9.8	Diatopie des alternances <e>/Ø, PROPERS CR2sg __#V	352
9.9	Diatopie des alternances <e>/Ø, 1100-1200, PROPERS CR2sg	353
9.10	Diatopie des alternances <e>/Ø, 1201-1322, PROPERS CR2sg	354
9.11	Distribution contextuelle des PROPERS CR1sg en défaut de <e>	356
9.12	<i>Roland</i> , ms. d'Oxford, fol. 1r, l. 21 (v. 21)	358
9.13	Diatopie des alternances <e>/Ø, PROPERS CR1sg __#C	360
9.14	Diatopie des alternances <e>/Ø, PROPERS CR1sg __#V	361

9.15 Diachronie des défauts de <e>, a.n. vs a.fr.cont., PROPERSCRS1sg	362
9.16 Diachronie des défauts de <e>, a.n. vs a.fr.cont, PROPERSCRS1sg __#V	363
9.17 Diatopie des alternances <e>/Ø, 1100-1200, PROPERSCRS1sg __#V	365
9.18 Diatopie des alternances <e>/Ø, 1201-1322, PROPERSCRS1sg __#V	366
9.19 Distribution contextuelle des PROPERSCRS1sg en défaut de <e>	368
9.20 Diatopie des alternances <e>/Ø, PROPERSCRS 1sg __#V	371
9.21 Diatopie des alternances <e>/Ø, 1100-1200, PROPERSCRS1sg __#V	372
9.22 Diatopie des alternances <e>/Ø, 1201-1322, PROPERSCRS1sg __#V	373
9.23 Distribution contextuelle des PROPERSCRS2sg en défaut de <e>	376
9.24 Distribution contextuelle des PROPERSFÉMCS3sg en défaut de <e>	378
9.25 Diatopie des alternances <e>/Ø PROPERSFÉMCS3sg	380
10.1 <i>Roland</i> , ms. d'Oxford, fol. 31r, l. 15 (v. 1698)	392
10.2 <i>Roland</i> , ms. d'Oxford, fol. 66v, l. 26 (v. 3698)	394
10.3 Diachronie des taux de <i>come</i> bisyllabiques devant PROPERSCRS et VER conjugué	395
10.4 Distribution contextuelle des CONJ <i>base</i> + <i>-que</i> en défaut de <e>	402
10.5 Diatopie des alternances <e>/Ø, CONJ en <i>base</i> + <i>-que</i> __#V	403
10.6 Diatopie des alternances <e>/Ø, CONJ __#V (nb. de textes / région > 5)	404
10.7 Diatopie des alternances <e>/Ø, 1100-1200, CONJ en <i>base</i> + <i>-que</i>	405
10.8 Diatopie des alternances <e>/Ø, 1201-1322, CONJ en <i>base</i> + <i>-que</i>	406
10.9 Diachronie des défauts de <e>, a.n. vs a.fr.cont, CONJ en <i>base</i> + <i>-que</i> __#V	407
10.10 Diachronie des défauts de <e>, a.n. vs a.fr.cont., CONJ en <i>base</i> + <i>-que</i> __#V	408
10.11 Distribution contextuelle des CONJ <i>que</i> en défaut de <V>	410
10.12 Diatopie des alternances <e>/Ø, élément <i>que</i> __#V	411
10.13 Diatopie des alternances <e>/Ø, élément <i>que</i> __#V (nb. de textes / région > 5)	412
10.14 Diachronie des défauts de <e>, a.n. vs a.fr.cont., élément <i>que</i> __#V	413
10.15 Diachronie des défauts de <e>, a.n. vs a.fr.cont, élément <i>que</i> __#V	413
10.16 Diatopie des alternances <e>/Ø, 1100-1200, élément <i>que</i> __#V	415
10.17 Diatopie des alternances <e>/Ø, 1201-1322, élément <i>que</i> __#V	416
10.18 Distribution contextuelle des CONJ « <i>come</i> » en défaut de <e> devant PROPERSCRS ou VER	419
10.19 Diatopie des alternances <e>/Ø, 1151-1200, <i>com</i> devant PROPERSCRS et VERcjg en #V	421
10.20 Diatopie des alternances <e>/Ø, 1151-1200, <i>com</i> devant PROPERSCRS et VER conjugué en #C	422
10.21 Diatopie des alternances <e>/Ø, CONJ « <i>come</i> » #C et #V	423
10.22 Distribution contextuelle des CONJ « <i>come</i> » en défaut de <e>	426
10.23 Diatopie des alternances <e>/Ø, CONJ « <i>come</i> » __#{C, V}	428
10.24 Diatopie des alternances <e>/Ø, 1100-1200, CONJ « <i>come</i> » __#C	430
10.25 Diatopie des alternances <e>/Ø, 1201-1322, CONJ « <i>come</i> » __#C	431
10.26 Diatopie des alternances <e>/Ø, 1100-1200, CONJ « <i>come</i> » __#V	433
10.27 Diatopie des alternances <e>/Ø, 1201-1322, CONJ « <i>come</i> » __#V	434
11.1 Distribution diachronique des occurrences de hiatus réduits dans le NCA.	454
11.2 Diatopie de la réduction de əV́	456
11.3 Diatopie de la réduction de əV́, 1100-1175	458
11.4 Diatopie de la réduction de əV́, 1176-1250	459
11.5 Diatopie de la réduction de əV́, 1251-1322	460
11.6 Distribution diachronique des occurrences de hiatus réduits dans le NCA.	468
11.7 Diatopie de la syncope	470
11.8 Diatopie de la syncope, 1100-1176 et 1176-1250	472
11.9 Diatopie de la syncope, 1251-1322	473
12.1 Taux de défaut de voyelles finale atone par catégorie	485
12.2 Taux de défaut de voyelles finale atone par catégorie et selon le contexte	486
12.3 Valeur- <i>t</i> (effet du contexte) pour chaque catégorie	488
12.4 Valeur- <i>t</i> (effet du contexte) pour chaque catégorie	490

12.5 Distribution des taux d'élision en fonction des catégories et items.	495
12.6 Distribution des taux d'élision en fonction de COND CV.	502
13.1 Représentation 3D du modèle modulaire d'association entre texte, mètre et événements musicaux	516
13.2 Représentation 3D d'association entre texte, mètre et événements musicaux pour une coupe épique	517
13.3 <i>Aucassin et Nicolette</i> , v. 1-2 (vii, iii, v), ms. et transcription (terminaison féminine)	518
13.4 « En un vergier », v. 1 et 3. ms. et transcription (élision)	522
13.5 « Bele Doette », v. 1 et 3, ms. et transcription (coupe épique)	525
13.6 « En un vergier », v. 1 et 3, ms. et transcription (coupe épique)	526
13.7 « La douce voiz... », v. 7, ms. et transcription (coupe épique)	527
14.1 Distribution des taux d'élision en fonction des catégories et items.	539
14.2 Diatopie des alternances <e>/Ø, CONJ en <i>base</i> + <i>-que</i> __#V	561
14.3 Distribution des taux d'élision en fonction des critères relevés.	567
14.4 Extrait de la fig. 5.10.b : taux d'élision des ADJ FÉMSG	570
14.5 Extrait de la fig. 9.19.b : taux d'élision du PRO PERS CSsg <i>jo</i>	571
E.1 <i>Chanson de Roland</i> , v. 2, dans le format du fichier <i>rolandox.xml</i> du NCA importé dans TXM.	610
E.2 Diachronie des noms en défaut de <e> __#V, en fonction des lemmes	611
E.3 Diachronie des noms en défaut de <e> __#C, en fonction des lemmes	612
E.4 Diatopie de la prévalence de l'élision, NOM	613
E.5 Distribution contextuelle des PRÉP « de » en défaut de <e>	614
E.6 Diatopie des formes picardes <i>le</i> du PRO PERS FÉM CR3sg (<i>vs la</i>) (DEES 1980).	615
E.7 Diatopie des formes en <i> initial du PRO PERS FÉM CS3sg (<i>vs la</i>) (DEES 1980).	616
E.8 Diatopie des alternances <e>/Ø, 1100-1200, PRO PERS MASC CR3sg __#V	617
E.9 Diatopie des alternances <e>/Ø, 1201-1322, PRO PERS MASC CR3sg __#V	618
E.10 Diatopie des alternances <e>/Ø, 1100-1200, PRO PERS MASC CR3sg __#C	619
E.11 Diatopie des alternances <e>/Ø, 1201-1322, PRO PERS MASC CR3sg __#C	620
E.12 Diatopie des alternances <e>/Ø, 1100-1200, PRO PERS CR1sg __#C	621
E.13 Diatopie des alternances <e>/Ø, 1201-1322, PRO PERS CR1sg __#C	622
E.14 Diatopie des alternances <e>/Ø, 1100-1200, PRO PERS CS3sg __#C	623
E.15 Diatopie des alternances <e>/Ø, 1201-1322, PRO PERS CS3sg __#C	624
E.16 Diatopie des alternances <e>/Ø, 1100-1200, PRO PERS CS3sg __#V	625
E.17 Diatopie des alternances <e>/Ø, 1201-1322, PRO PERS CS3sg __#V	626
E.18 Diatopie des alternances <V>/Ø, 1100-1200, ART DÉF MASC CSsg __#V	627
E.19 Diatopie des alternances <V>/Ø, 1100-1200, ART DÉF MASC CSsg __#V	628
E.20 Diatopie des alternances <e>/Ø, 1100-1200, CONJ « come » __#C	629
E.21 Diatopie des alternances <e>/Ø, 1201-1322, CONJ « come » __#C	630
E.22 Diatopie des alternances <e>/Ø, 1100-1200, CONJ « come » __#V	631
E.23 Diatopie des alternances <e>/Ø, 1201-1322, CONJ « come » __#V	632
E.24 Distribution des longueurs de vers détectées par le PAM dans <i>Alexis</i>	633
E.25 Distribution des longueurs de vers détectées par le PAM dans <i>Roland</i>	634
E.26 Mesures exploratoires pour détecter un lien entre schwa final et vers mal formés dans <i>Roland</i>	635

Liste des tableaux

2.1	Métadonnées des textes des catégories (a) et (b).	21
2.2	NCA : qualité des éditions	23
2.3	Combinaisons des métadonnées du NCA3	24
2.4	NCA : obtention d'une <i>regionDees</i> à partir des localisations philologiques	32
2.5	NCA : régions trop affaiblies par une restriction de qualité éditoriale	40
2.6	BFM : nombre de textes par dialecte	43
3.1	Dix premières lignes de R ADJ <i>femi_sg merged freq.xlsx</i>	70
3.2	Dix premières lignes de R ADJ <i>femi_sg merged freq.xlsx</i> avec tri	72
3.3	Dix premières lignes de ADJ <i>irrelevant forms.xlsx</i>	73
3.4	Données récapitulatives non contextuelles, ADJ FÉMSG	80
3.5	Données récapitulatives contextuelles, ADJ FÉMSG	80
3.6	Ensemble des sorties graphiques, cartographiques et tabulaires de Comparalem	89
3.7	Sortie tabulaire de <i>pamviz.R</i> pour <i>Roland</i>	106
4.1	Quantité de noms, de lemmes à variation, et d'occurrences en défaut de <e>.	123
4.2	NCA : Lemmes de noms à défaut de Ø	124
5.1	Quantité d'adjectifs, de lemmes à variation, et d'occurrences en défaut de <e>.	136
5.2	Fréquence et taux de défaut de <e> par lemme, ADJ FÉMSG	147
5.3	Taux de défaut de <e> __#V en anglo-normand par période	173
5.4	Taux de défaut de <e> __#V en anglo-normand de période 3 (1201-1250)	173
5.5	Taux de défaut de <e> __#V en anglo-normand en période 4 (1251-1322)	174
6.1	Quantité de prépositions, de lemmes à variation, et d'occurrences en défaut de <e>.	185
6.2	Fréquence et taux de défaut de <e> par lemme, PRÉP	200
6.3	Taux de défaut de <e> __#V en anglo-normand par période, ADJ FÉMSG vs PRÉP « de »	229
7.1	Combinaisons des formes et étiquettes de pos pour les formes <i>ne</i> , <i>ni</i> et <i>n</i>	252
7.2	Quantité d'adverbes, de lemmes à variation, et d'occurrences en défaut de <e>.	254
7.3	Quantité d'adverbes, de lemmes à variation, et d'occurrences en défaut de <e>, sans « ensemble ».	254
7.4	Fréquence des lemmes se terminant en <e> (hors <que>), ADV	255
7.5	Fréquence et taux de défaut de <e> par lemme, ADV	266
7.6	Fréquence et taux de défaut de <e> par lemme et par contexte, ADV	271
7.7	Taux de défaut de <e> __#V en anglo-normand par période, ADJ FÉMSG vs PRÉP « de » vs ADV	296
8.1	Quantité de formes et taux de formes en défaut de <e> pour les articles	303
9.1	Paradigme des pronoms personnels	328
9.2	Quantité de formes et taux de formes en défaut de <e> pour les pronoms personnels	333
9.3	Enclises du PROPERSMASC3sg	341
10.1	CONJ sujettes au défaut de <e> (F > 200)	389
10.2	Formes et fréquence des formes de l'élément <i>que</i> dans le NCA	397
10.3	Formes et fréquence des formes de « come » dans le NCA	397
10.4	Quantité de formes et taux de formes en défaut de <e> pour les conjonctions	399
10.5	Superposition variable des taux moyens régionaux les plus bas de <i>com</i>	429
11.1	Premières occurrences de formes à hiatus əV réduits dans le NCA.	453
11.2	Premières occurrences de formes syncopées dans le NCA.	466

12.1 Synthèse des défauts de <e> ou <V> par catégorie	483
-----------------------------------------------------------------	-----

Table des matières

Sommaire	5
1 Introduction	1
1.1 Objet de ce travail	1
1.1.1 Alternance des voyelles finales atones avec zéro	1
1.1.1.1 Distinguer l'élision de l'apocope et de l'épenthèse	3
1.1.1.2 Origine des finales atones régulières de l'a.fr.	3
1.1.2 Projet de cette thèse	4
1.2 Structures et conventions de ce document	5
1.2.1 Structure	5
1.2.2 Conventions	6
1.2.2.1 Abréviations bibliographiques	7
I CORPUS ET OUTILS	8
2 Corpus	10
2.1 Le <i>Nouveau Corpus d'Amsterdam</i>	10
2.1.1 La genèse du NCA : l' <i>Atlas</i> de DEES (1987)	13
2.1.2 De l' <i>Atlas</i> au NCA	18
2.1.3 État des métadonnées	20
2.1.4 Décomposition du corpus en fonction des métadonnées	22
2.1.4.1 Métadonnées : la place de l' <i>Atlas</i> de 1987 dans le NCA	23
2.1.4.2 Distribution diachronique des textes	25
2.1.4.3 Distribution diatopique des textes	27
2.1.5 Choix des métadonnées utilisées	29
2.1.5.1 Métadonnées diachroniques	29
2.1.5.2 Métadonnées diatopiques	30
1100-1150.	33
1151-1200.	33
1201-1250.	38
1251-1322.	38
2.1.5.3 Métadonnées qualitatives	39
2.1.6 En somme : le NCA ici utilisé	40
2.2 BFM : la <i>Base de Français Médiéval</i>	41
2.2.0.1 Composition du corpus	42
2.2.0.2 Structure de l'annotation	43
2.2.1 Comparaison entre NCA et BFM	45
2.3 Sources musicales : philologie musicale des XII ^e et XIII ^e s.	48
2.3.1 Chansons de trouvères, <i>Aucassin et Nicolette</i> et <i>Li Gieus de Robin et de Marion</i>	48
2.3.1.1 Les témoins des trouvères	49
Le ms. U.	49
Les ms. M, N, P, X et K.	50
Le ms. V.	50
Le ms. O.	50
Le ms. a.	50
Le ms. i.	50
2.3.1.2 <i>Aucassin et Nicolette</i>	51
2.3.1.3 <i>Li Gieus de Robin et Marion</i>	52

2.3.1.4	La question rythmique	53
2.3.1.5	Musique notée et itération musicale	59
2.3.1.6	En somme	62
2.3.2	En somme	62
2.4	Conclusion	63
3	Outils	64
3.1	TXM	64
3.1.1	Requêtes CQP de type Index	64
3.1.2	En somme	66
3.2	Comparalem	67
3.2.1	Flux de travail de Comparalem	68
3.2.1.1	Étape préliminaire : préparation des données	68
En somme.	69	
3.2.1.2	Première étape : générer la liste des lemmes à variation	69
Du point de vue de l'utilisateur.	70	
Intervention de l'utilisateur sur [POS]_merged_freq.	71	
Intervention de l'utilisateur : [POS]_irrelevant_forms.	72	
3.2.1.3	Seconde étape : observer la distribution des variantes	72
Import des données.	73	
Filtrage en fonction des lemmes.	74	
Filtrage en fonction des occurrences exclues.	74	
Filtrage en fonction des dates.	74	
Filtrage en fonction de la flexion.	75	
Production de sorties graphiques diachroniques.	75	
Production de sorties graphiques par lemme.	75	
Production de données quantitatives.	80	
Production de sorties cartographiques synchroniques.	80	
Production de sorties cartographiques diachroniques.	81	
Export des sorties graphiques et quantitatives.	88	
Du point de vue de l'utilisateur.	88	
Synthèse du flux de travail de la seconde étape.	89	
3.2.2	Note sur la reproductibilité et la diffusion du programme	91
3.2.3	En somme	91
3.3	Le PAM, <i>Programme d'Analyse Métrique</i>	91
3.3.1	Flux de travail	92
3.3.1.1	Préparation des textes : pam_corpus	92
xml_txt.py.	93	
clear_numbers.py.	93	
3.3.1.2	Exécution du PAM	93
syllabise.py.	94	
Annotation prosodique.	94	
Analyse métrique de premier niveau.	97	
Analyse métrique de second niveau.	98	
Affichage des résultats.	100	
Exports.	103	
Du point de vue de l'utilisateur.	103	
3.3.2	Configurations linguistiques spécifiques : les <i>config</i>	104
3.3.3	Post-traitement des résultats en R	105
3.3.3.1	Rectification de m	105
3.3.3.2	Distribution des nombres de syllabes métriques	106
3.3.3.3	Position des vers dans le texte	107
3.3.3.4	Relation entre schwas et longueur du vers	107
A. Quantité de schwas par vers.	108	

B. Quantité de schwas en fonction de la longueur des vers.	108
C. Présence de schwas en fonction du nombre de vers.	110
D. Vue d'ensemble.	110
3.3.3.5 En somme	110
3.3.4 Note sur la reproductibilité et la diffusion du programme	110
Le PAM.	111
Syllabize.py	111
Pam_corpus.	111
pamviz.R.	111
3.3.5 En somme	111
3.4 Conclusion	113

II DÉPOUILLEMENT DES ALTERNANCES GRAPHIQUES 115

4 Noms communs	117
4.1 Introduction	117
4.1.1 Rappel de méthode : Comparalem	117
4.1.1.1 Requêtes CQP	118
4.1.1.2 Traitement des jeux de données dans Comparalem	119
4.1.2 Exclusions	120
4.2 Présentation des résultats	123
4.2.1 Typologie	124
4.2.2 Distributions	125
4.3 Généralisations	131
4.4 Conclusion	132
5 Adjectifs au féminin singulier	133
5.1 Introduction	133
5.1.1 Choix des adjectifs au féminin singulier	133
5.1.2 Exclusions	135
5.1.3 Relemmatisation	136
5.2 Présentation des résultats	136
5.2.1 Typologie	137
5.2.2 Distribution diachronique générale	138
5.2.3 Distribution des lemmes en défaut de <e> : diachronie et fréquence	139
5.2.3.1 Fréquence des lemmes et défaut de <e>	146
5.2.3.2 Fréquence, collocations et défauts de <e>	148
<i>Saint(e) + nom</i>	148
<i>Long(e) + nom</i>	149
<i>mal + nom</i>	149
Autres lemmes et défaut de méthode	152
5.2.4 Distribution contextuelle	153
5.2.5 Distribution diatopique	155
5.2.6 Interface des distributions diatopiques et diachroniques	158
1100 - 1150	159
1151-1200	161
1201-1250	163
1251-1322	165
5.3 Des observations aux généralisations	167
5.3.1 Lemmes atteints par le défaut de <e>	167
5.3.2 Le rôle du contexte	170
5.3.3 Diachronie générale	171
5.3.4 Centre - périphérie	171

5.3.5	Poids de l'anglo-normand et de son hétérogénéité	172
5.3.6	Place de la variation au sein des ADJFÉM _{sg}	177
5.4	Conclusion	178
6	Prépositions	180
6.1	Introduction	180
6.1.1	Exclusions	181
6.1.2	Relemmatisation	184
6.2	Présentation des résultats	185
6.2.1	Typologie	186
6.2.2	Distribution diachronique générale	187
6.2.3	Distribution des lemmes en défaut de <e> : diachronie et fréquence	188
6.2.3.1	Lemmes attestés en défaut de <e> dès <i>Roland</i>	188
	<i>D</i> (« de »)	188
	<i>Entr</i>	193
	<i>Sor</i>	194
6.2.3.2	Lemmes attestés en défaut de <e> au cours du xii ^e s.	195
	<i>Encontr</i>	195
	<i>Contr</i>	196
	<i>Desor</i>	196
	Un lemme, plusieurs attributs de POS	197
	En somme	199
6.2.3.3	Fréquence des lemmes et défauts de <e>, ou plutôt typologie des défauts	200
6.2.3.4	Fréquence, collocations et défauts de <e>	201
	<i>Entr</i>	201
	<i>Contr</i>	202
	<i>Encontr</i>	204
	En somme	205
6.2.4	Distribution contextuelle	205
6.2.4.1	« De », « entre », « contre » et « rencontre »	208
	« De »	208
	« Entre »	211
	« Contre » et « rencontre »	213
6.2.4.2	« Sor » et « desor »	215
6.2.4.3	En somme	215
6.2.5	Distribution diatopique	217
6.2.5.1	« De »	220
6.2.5.2	« Entre »	222
6.2.5.3	« Sor » et « desor »	224
6.2.5.4	En somme	226
6.2.6	Interface des distributions diatopiques et diachroniques	226
6.2.6.1	« De »	226
	Domaine continental	226
	Angleterre	229
6.2.6.2	« entre »	230
6.2.6.3	En somme	233
6.3	Des observations aux généralisations	233
6.3.1	Lemmes atteints par le défaut de <e>	233
6.3.2	Le rôle du contexte	236
6.3.3	Diachronie générale	238
6.3.4	Diatopie	239
6.3.5	Poids de l'anglo-normand et de son hétérogénéité	239
6.3.6	Place de la variation au sein des PRÉP	244
6.4	Conclusion	245

7	Adverbes	247
7.1	Introduction	247
7.1.1	Note sur la pérennité des alternances <e>/∅ (et <es>) des adverbes	247
7.1.2	Exclusions	249
7.1.2.1	Úsque : motivation de l'exclusion	250
7.1.2.2	Ne : motivation de l'exclusion	251
7.1.2.3	En somme	252
7.1.3	Relemmatisation	253
7.2	Présentation des résultats	253
7.2.1	Typologie	254
7.2.2	Distribution diachronique générale	255
7.2.3	Distribution des lemmes en défaut de <e> : diachronie et fréquence	258
7.2.3.1	Lemmes attestés en défaut de <e> dès <i>Roland</i>	259
	Spécificité de <i>ensembl</i>	262
	Annotation en POS : adverbe ou préposition ?	262
7.2.3.2	Lemmes attestés en défaut de <e> au cours du xiii ^e s.	263
	<i>Lors</i>	263
	<i>Encor</i>	263
	<i>Derrier</i>	264
	<i>Arrier</i>	264
	— Spécificité de « derriere » et « arriere »	265
	En somme	265
7.2.3.3	Fréquence des lemmes et défauts de <e>	266
7.2.3.4	Fréquence, collocations et défauts de <e>	267
	<i>Ensembl od</i>	267
7.2.4	Distribution contextuelle	268
7.2.4.1	En somme	272
7.2.5	Distribution diatopique	272
7.2.5.1	En somme	278
7.2.6	Interface des distributions diatopiques et diachroniques	278
7.2.6.1	1100-1150	279
	__#V	279
	Prévalence de l'élision et __#C	280
7.2.6.2	1151-1200	281
	__#V	281
	Prévalence de l'élision et __#C	282
7.2.6.3	1201-1250	284
	__#V	284
	Prévalence de l'élision et __#C	285
7.2.6.4	1251-1322	286
	__#V	286
	Prévalence de l'élision et __#C	287
7.2.6.5	En somme	288
7.3	Des observations aux généralisations	288
7.3.1	Lemmes atteints par le défaut de <e>	288
7.3.2	Le rôle du contexte	291
7.3.3	Diachronie générale	293
7.3.4	Diatopie	294
7.3.5	Poids de l'anglo-normand et de son hétérogénéité	295
7.3.6	Place de la variation au sein des Adv	297
7.4	Conclusion	298

8	Articles	300
8.1	Introduction	300
8.1.1	Les articles de l'ancien français	300
8.1.2	Traitement des données	302
8.2	Présentation des résultats	302
8.2.1	Typologie	303
8.2.2	Articles à élision systématique : <i>le, la</i>	304
8.2.2.1	Défauts devant #V	306
8.2.2.2	Défauts devant #C	309
8.2.2.3	En somme	311
8.2.3	Articles à élision facultative : <i>li</i> (CSsg), <i>une</i>	311
8.2.3.1	Défauts devant #V	311
	<i>L</i> MASC CSsg	311
	<i>Un</i> FÉMSG	314
8.2.3.2	Défauts devant #C	316
	<i>L</i> MASC CSsg	316
	<i>Un</i> FÉMSG	316
8.2.3.3	Distribution diatopique	317
	<i>L</i> MASC CSsg	318
	<i>Un</i> FÉMSG	319
8.2.3.4	En somme	321
8.2.4	Article à élision prohibée : <i>li</i> (CSpl)	321
8.3	Des observations aux généralisations	323
8.3.1	Lemmes atteints par l'enclise et l'élision	323
8.3.2	Le rôle du contexte	324
8.3.3	Diachronie et diatopie	325
8.4	Conclusion	325
9	Pronoms personnels	327
9.1	Introduction	327
9.1.1	Les pronoms personnels de l'ancien français	327
9.1.1.1	<i>Je</i> CS : progression de l'élision	328
	<i>Tu</i> (<i>te</i>), chemin parallèle ?	330
9.1.1.2	Pronoms régimes	331
9.1.1.3	Circularité de l'élision et de l'identification des formes faibles, et nature du phénomène de défaut	331
	En somme	332
9.1.2	Traitement des données	332
9.2	Présentation des résultats	333
9.2.1	Trouble dans le genre des PRO PERS CR3sg	334
9.2.2	Typologie	335
9.2.2.1	Défauts devant #C	338
	Enclise	338
	<i>Ele</i> ~ <i>el</i>	340
9.2.2.2	En somme	340
9.2.3	PRO PERS MASC CR3sg : <i>le</i> ~ <i>l</i>	341
9.2.3.1	Distribution contextuelle et diachronique	341
	<i>l</i> __#C	341
	<i>l</i> __#V	343
9.2.3.2	Distribution diatopique	344
9.2.3.3	Intersection de la diatopie et de la diachronie	346
9.2.3.4	En somme	346

9.2.4	PROPERSCR2sg : <i>te</i> ~ <i>t</i>	346
9.2.4.1	Distribution contextuelle et diachronique	346
	<i>t</i> __#C	346
	<i>t</i> __#V	348
9.2.4.2	Distribution diatopique	351
	<i>t</i> __#C	351
	<i>t</i> __#V	352
9.2.4.3	Intersection de la diatopie et de la diachronie	352
9.2.4.4	En somme	355
9.2.5	PROPERSCR1sg : <i>me</i> ~ <i>m</i>	355
9.2.5.1	Distribution contextuelle et diachronique	355
	<i>m</i> __#C	357
	<i>m</i> __#V	358
9.2.5.2	Distribution diatopique	360
	<i>m</i> __#C	360
	<i>m</i> __#V	361
9.2.5.3	Intersection de la diatopie et de la diachronie	363
	<i>m</i> __#C	363
	<i>m</i> __#V	364
9.2.5.4	En somme	364
9.2.6	PROPERSCS1sg : <i>je, jou</i> (etc.) ~ <i>j</i>	367
9.2.6.1	Distribution contextuelle et diachronique	367
	<i>j</i> __#C	369
	<i>j</i> __#V	369
9.2.6.2	Distribution diatopique	370
9.2.6.3	Intersection de la diatopie et de la diachronie	371
9.2.6.4	En somme	374
9.2.7	PROPERSCS2sg : <i>tu, te</i> ~ <i>t</i>	374
9.2.7.1	Distribution contextuelle et diachronique	375
	<i>t</i> __#C ?	375
	<i>t</i> __#V	375
9.2.7.2	En somme	377
9.2.8	PROPERSFÉMCS3sg : <i>ele</i> ~ <i>el</i>	377
9.2.8.1	Distribution contextuelle et diachronique	377
	<i>el</i> __#C	377
	<i>el</i> __#V	377
9.2.8.2	Distribution diatopique	379
9.2.8.3	Intersection de la diatopie et de la diachronie	379
9.2.8.4	En somme	379
9.3	Des observations aux généralisations	381
9.3.1	Lemmes atteints par l'enclise et l'élision	381
9.3.2	Le rôle du contexte	383
9.3.3	Diachronie générale	384
9.3.4	Diatopie générale	385
9.4	Conclusion	385
10	Conjonctions : <i>Que, -que</i> et <i>come</i>	388
10.1	Introduction	388
10.1.1	Les conjonctions de l'ancien français	388
10.1.2	<i>Que</i> , une question de nature	389
10.1.3	<i>Come</i> , existence puis disparition d'une distribution syntaxique	391
	10.1.3.1 Pertinence d'une distribution syntaxique	391
	10.1.3.2 Affaissement de la distribution syntaxique	394

10.1.4	Traitement des données	396
10.1.4.1	Requêtes pour l'élément <i>que</i> et pour <i>com(e)</i>	396
10.1.4.2	Relemmatisation	397
10.2	Présentation des résultats	398
10.2.1	Typologie	399
10.2.2	Locutions en <i>base + -que</i>	400
10.2.2.1	Distribution diachronique et contextuelle	401
10.2.2.2	Distribution diatopique	403
10.2.2.3	Intersection de la diatopie et de la diachronie	404
10.2.2.4	En somme	408
10.2.3	L'élément <i>que</i>	409
10.2.3.1	Distribution diachronique et contextuelle	409
10.2.3.2	Distribution diatopique	411
10.2.3.3	Intersection de la diatopie et de la diachronie	414
10.2.3.4	En somme	414
10.2.4	<i>Com(e)</i>	417
10.2.4.1	Alternance <i>com</i> ~ <i>come</i> en contexte syntaxique de <i>com</i>	417
	Distribution diachronique et contextuelle de la neutralisation	418
	Début de la neutralisation	420
	Distribution diatopique de la neutralisation	421
	En somme	424
10.2.4.2	Alternances <i>com</i> ~ <i>come</i> dans tous les contextes	425
	Distribution diachronique et contextuelle	425
	Distribution diatopique	427
	Intersection de la diatopie et de la diachronie	429
	En somme	432
10.2.4.3	En somme	435
10.3	Des observations aux généralisations	435
10.3.1	Lemmes atteints par le défaut de <e>	436
10.3.2	Le rôle du contexte	437
10.3.3	Diachronie générale	437
10.3.4	Diatopie	438
10.3.5	Place de la variation au sein des CONJ	438
10.4	Conclusion	439
11	Chutes des schwas prétoniques	442
11.1	Introduction	442
11.1.1	Origine et destin des schwas prétoniques	443
11.1.1.1	Schwa interne en hiatus	444
11.1.1.2	Schwa prétonique hors hiatus	448
11.2	Développement des chutes de schwa en hiatus əV̆	450
11.2.1	Diachronie de la réduction des hiatus əV̆	451
11.2.1.1	Méthode : sélection de huit lemmes hiatus	452
11.2.1.2	Premières attestations de la réduction des hiatus əV̆	452
11.2.1.3	Cooccurrences des graphies <eV> et <V>	453
11.2.1.4	En somme	455
11.2.2	Diatopie de la réduction des hiatus əV̆	455
11.2.3	Intersection de la diachronie et de la diatopie	457
11.2.3.1	1100-1175	458
11.2.3.2	1176-1250	459
11.2.3.3	1251-1322	460
11.2.3.4	En somme	461

11.3	Chute de schwa internes hors hiatus	461
11.3.1	Diachronie de la réduction des schwas hors hiatus	462
11.3.1.1	Méthode	463
11.3.1.2	Premières attestations de la syncope	464
11.3.1.3	Cooccurrences des graphies syncopées et non syncopées	467
11.3.2	Diatopie de la syncope	469
11.3.2.1	En somme	471
11.3.3	Intersection entre la diachronie et la diatopie	471
11.3.3.1	1100-1175	471
11.3.3.2	1176-1250	473
11.3.3.3	1251-1322	474
11.3.3.4	En somme	474
11.4	Discussion	475
11.4.1	Synthèse	475
11.4.1.1	Réduction des hiatus	475
11.4.1.2	Syncope	476
11.4.2	Points communs et différences	476
11.4.2.1	La coexistence des graphies en <e> et sans <e> est commune	476
11.4.2.2	La motivation phonologique ne peut être commune	477
11.4.2.3	Comparaison avec les schwas finaux	478
11.4.3	En somme	478
11.5	Conclusion	479

III ANALYSE 480

12	Généralisations sur les données graphiques	482
12.1	Synthèse des taux de défaut	482
12.1.1	Typologie selon les taux absolus de défaut	483
12.1.2	Typologie selon les taux contextuels de défaut	485
12.1.3	Typologie selon l'effet du contexte	487
12.1.4	En somme	487
12.2	Ce qui tient de l'élision, et ce qui n'en tient pas	488
12.2.1	Interface entre valeurs- <i>t</i> et taux contextuels	489
12.2.2	Interprétation linguistique des valeurs- <i>t</i> , et intégration d'autres variables	491
12.3	Généralisations morphosyntaxiques et phonologiques	496
12.3.1	Élimination de la condition morphosyntaxique : COND CV	497
12.3.1.1	COND CV et les pronoms sujets de 1sg et 2sg	497
	<i>Je, jo</i>	498
	<i>Tu, te</i>	498
12.3.1.2	COND CV, <i>li</i> sg élidable et <i>li</i> pl non élidable	499
	<i>Li</i> pl	499
	<i>Li</i> sg	499
	Retour à <i>Tu, te</i>	500
12.3.1.3	En somme	501
12.3.2	Les collocations	501
12.3.2.1	<i>Ensembl od</i> , parangon de la collocation	502
12.3.2.2	<i>Entr eus</i> et <i>(en)contr eus</i>	503
	« Entre »	503
	« (En)contre »	504
12.3.2.3	Collocation des adjectifs	505
12.3.2.4	En somme	506
12.4	Conclusion	507

13	L'élision en métrique et musication	509
13.1	Introduction	509
13.1.1	Métrique générative	510
13.1.2	Patron métrique, grammaire d'association et extramétricalité	510
13.1.3	Application à l'ancien français	512
13.1.3.1	Patrons métriques	512
13.1.3.2	Grammaire d'association et extramétricalité	513
13.1.3.3	Extramétricalité et musication	516
13.1.4	Échec heuristique du PAM	519
13.2	L'élision métrique	520
13.2.1	L'élision métrique médiévale	520
13.2.1.1	Élision métrique et musication	522
13.2.2	Élision métrique et coupe épique	522
13.2.2.1	Coupe épique devant initiale consonantique	523
13.2.2.2	Élision de coupe épique devant initiale vocalique	524
	Modification de la constituance neumatique	524
	Duplication	526
	Ambiguïté entre les sources	526
13.2.2.3	En somme	528
13.2.3	Domaine où l'élision métrique a lieu	528
13.3	Naturalité du domaine de l'élision métrique, paraphonologie et conflit	529
13.3.1	Congruence entre l'élision métrique et celle du xvi ^e s.	530
13.3.1.1	En somme	532
13.3.2	Paraphonologie métrique et alignement des domaines	533
13.3.3	En somme	533
13.4	Conflit de domaine	534
13.4.1	Élargissement du domaine computationnel	534
13.4.2	Conséquences de l'élargissement du domaine computationnel et figement graphique	535
13.5	Conclusion	536
14	Analyse phonologique	538
14.1	Introduction	538
14.1.1	Résumé des données	539
14.1.1.1	Période 1	539
14.1.1.2	Période 2	540
14.1.2	Cahier des charges général	541
14.2	Condition phonologique : le flottement	541
14.2.1	Flottement diacritique et diachronique	543
14.2.1.1	Transcription du diacritique par le flottement	543
14.2.1.2	Origine du flottement diacritique	545
14.2.1.3	En somme	547
14.2.2	Flottement non diacritique et synchronique	547
14.2.2.1	Les hiatus et schwas internes	548
14.2.2.2	En somme	548
14.2.3	Mécanisme phonologique	549
14.2.3.1	Tranel : contrainte anti-hiatique, ou contrainte anti-mélodie flottante	549
14.2.3.2	Flottement et élision en fr.mod. (1) : ə sans V	550
14.2.3.3	Flottement et élision en fr.mod. (2) : ə ⇔ V	552
14.2.3.4	Mécanisme retenu pour l'a.fr. : ə ⇏ V	553
	Externalité positive de *VCVIDES (1) : hiatus internes.	555
	*VCVIDES n'est pas un mécanisme anti-hiatique.	555
14.2.3.5	Note sur <i>que</i>	557
14.2.4	En somme	559

14.3	Domaines de l'élision	560
14.3.1	Période 1 : élision des clitiques	560
14.3.1.1	Condition de proximité syntaxique	562
14.3.1.2	Condition de taille	564
14.3.1.3	En somme	565
14.3.2	Période 2 : élision au sein d'un domaine plus large	565
14.3.2.1	Un potentiel candidat : le syntagme phonologique ϕ	566
	Mécanismes de restructuration de ϕ	568
	Problème heuristique de la restructuration de ϕ	568
14.3.2.2	Un candidat fiable, mais insuffisant : la syntaxe	569
	Quelques adjectifs élidés.	569
	Le chemin de <i>jo</i> , une fenêtre sur l'élargissement du domaine.	570
	En somme.	572
14.3.3	Fonctionnement phonologique synchronique	572
14.3.4	En somme	575
14.4	Conclusion	575
15	Conclusion	577
15.1	Résumé des résultats et analyses	577
15.2	Apports	578
15.2.1	Apport théorique : la réduction des hiatus, conséquence et non cause	578
15.2.2	Apports en phonétique historique	579
15.2.2.1	L'élision et l'élargissement du domaine computationnel	579
15.2.2.2	Flottement et clitiques	579
15.2.3	Apports techniques	580
15.3	Ouvertures pour parfaire ce travail	580
	ANNEXES	582
A	Abréviations, symboles, notations formelles	583
A.1	Abréviations et symboles	583
A.1.1	Périodisation	583
A.1.2	Abréviations générales	583
A.1.3	Abréviations d'étiquettes morphosyntaxiques	584
A.1.4	Dialectes	585
A.2	Notations	586
B	Liste des généralisations, règles, principes, etc.	587
C	Corpus informatiques : liste des textes	596
C.1	NCA : liste des textes	596
C.2	BFM : liste des textes cités	608
D	Corpus manuscrits : liste des témoins	609
D.1	Manuscrits des trouvères	609
D.2	Manuscrits d' <i>Aucassin et Nicolette</i> et du <i>Gieus de Robin et de Marion</i>	609
E	Figures rejetées en annexes	610
	Références	636
	Table des matières	652